

Copyright
by
Weilu Lin
2005

The Dissertation Committee for Weilu Lin
certifies that this is the approved version of the following dissertation:

**Closed-loop Subspace Identification and Fault Diagnosis
with Optimal Structured Residuals**

Committee:

S. Joe Qin, Supervisor

Thomas F. Edgar

Roger T. Bonnecaze

Glenn Y. Masada

Maruthi Ram Akella

**Closed-loop Subspace Identification and Fault Diagnosis
with Optimal Structured Residuals**

by

Weilu Lin, B.S., M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2005

Dedicated to my family.

Acknowledgments

I wish to thank the multitudes of people who helped me. The first one on the list is my supervisor Dr. S. Joe Qin, who patiently introduced me into the system identification and fault diagnosis areas. Without his guidance and support I will never finish this work at this point. I would like to thank Dr. L. Ljung for his valuable help and insight. Each of my committee members, Dr. Thomas F. Edgar, Dr. Roger T. Bonnecaze, Dr. Glenn Y. Masada, and Dr. Maruthi Ram Akella, has offered encouragement and support during the past five years.

I want to appreciate the help that I have received from my fellow graduate students in control groups at UT Austin. Thanks due to Dr. Jurgen Hahn, Dr. Henry Portrykus, Dr. Jin Wang, Dr. Richard Good, Qinghua He, Elaine Hale, Greg Cherry, Chris Harrison, Brent Bregenzer, Clare Schoene, Jie Yu, Quan Shen, and Kai Zhang. Special thanks to Mr. Koichi Onodera for frequent discussions and cooperations.

I would like also thank Dr. John Watkins and the management team at Weyerhaeuser who gave me valuable opportunities to work on real industrial problems. Thanks go to Dr. Christopher Alan McNabb, Dr. Anthony P. Swanda, Dr. Apostolos Rigopolous, Dr. Gary Wong, and Mr. Gary Gover. I would say that the internship at Weyerhaeuser is the most valuable time I

spent in the past five years.

Finally, I would like to thank to my elder brother, Weican Lin, who takes very good care of my parents when I am not there.

WEILU Lin

The University of Texas at Austin

May 2005

Closed-loop Subspace Identification and Fault Diagnosis with Optimal Structured Residuals

Publication No. _____

Weilu Lin, Ph.D.

The University of Texas at Austin, 2005

Supervisor: S. Joe Qin

The development of system identification and fault diagnosis theory is of great practical significance. Systems are concerned with a broad spectrum of human-made machinery, including industrial production facilities (power plants, chemical plants, oil refinery, semiconductor fabrication plants, steel mills, paper mills, etc.), transportation vehicles (ships, airplanes, automobiles) and household appliances (heating/air conditioning equipment, refrigerators, washing machines, etc.). This dissertation is focused on subspace identification algorithms and optimal structured residuals approach for processes modeling and diagnosis.

Main contributions of this work include:

1. Novel subspace identification methods (SIMs) with enforced causal models are implemented. It has been shown that proposed algorithm has

lower estimation variance compared to traditional SIMs. Meanwhile the rigorous analysis shows that the proposed algorithms are consistent under certain assumptions.

2. The feasibility of closed-loop subspace identification is investigated. Novel closed-loop subspace identification methods with innovation estimation are proposed. The new algorithms are shown to be consistent under closed-loop conditions, while the traditional SIMs fail to provide consistent estimates.
3. A new optimal structured residuals (OSR) approach for unidirectional fault diagnosis is proposed. The necessary and sufficient conditions for unidirectional fault isolability with OSR approach are introduced.
4. The OSR for unidirectional fault diagnosis is extended to multidimensional fault diagnosis. The sufficient condition for deterministic multidimensional fault isolability is investigated.

Table of Contents

Acknowledgments	v
Abstract	vii
List of Tables	xii
List of Figures	xiii
Chapter 1. Introduction and Dissertation Outline	1
Chapter 2. A Novel Subspace Identification Approach with Enforced Causal Models	6
2.1 Analysis of subspace formulation	7
2.1.1 Problem formulation and assumptions	7
2.1.2 Analysis of conventional SIMs	10
2.2 Subspace identification avoiding non-causal terms	13
2.2.1 PARSIM algorithms	14
2.2.2 Improved variance of PARSIM algorithms	18
2.2.3 Determination of observability matrix	22
2.3 Numerical implementation of PARSIMs	23
2.3.1 QR implementation for K	24
2.3.2 Determination of B, D	25
2.4 Simulation and industrial case studies	27
2.4.1 Simulation example 1	27
2.4.2 Simulation example 2	30
2.4.3 Industrial case study	32
2.5 Summary	33

Chapter 3. Closed-loop Subspace Identification with Innovation Estimation	35
3.1 Introduction	35
3.2 Analysis of subspace identification under closed-loop condition	37
3.2.1 Problem formulation and assumptions	37
3.2.2 Analysis of the closed-loop SIM	40
3.3 Closed-loop subspace identification methods with innovation estimation	44
3.3.1 Algorithm 1: PARSIM-E	44
3.3.2 Algorithm 2: PARSIM-E1	46
3.3.3 K estimation under closed-loop condition	49
3.3.4 Simulation studies	51
3.3.4.1 Simulation example: a SISO process	51
3.3.4.2 Simulation example: a MIMO process	52
3.4 Comparisons of closed-loop subspace identification methods . .	54
3.4.1 Closed-loop Subspace Identification Methods	55
3.4.1.1 Overview of closed-loop SIMs	55
3.4.1.2 Whitening Filter Approach	57
3.4.2 Simulation Study	59
3.5 Summary	60
 Chapter 4. An Optimal Structured Residual Approach for Unidirectional Faulty Sensor Diagnosis	 81
4.1 Introduction	81
4.2 Model and fault representation	84
4.3 Fault identification with optimal structured residuals	85
4.3.1 Review of structured residual approaches	86
4.3.2 The design of optimal structured residuals	87
4.3.3 Fault identification indices	90
4.4 Fault isolability	92
4.5 Reconstruction of normal measurements	96
4.6 An industrial boiler case study	97
4.7 Summary	103

Chapter 5. An Optimal Structured Residuals Approach for Multidimensional Faulty Sensor Diagnosis	105
5.1 Introduction	105
5.2 Fault diagnosis for dynamic systems	107
5.2.1 Problem formulation	107
5.2.2 Temporal redundancy and its relationship with dynamic PCA	111
5.3 Multidimensional faults diagnosis with optimal structured residuals	112
5.3.1 Optimal structured residuals approach	112
5.3.2 Fault identification indices	115
5.4 Fault isolability	117
5.5 Simulation example: a 4×4 dynamic process	120
5.6 Summary	123
Chapter 6. Conclusions and Future Research Suggestions	128
6.1 Contributions	128
6.2 Future research suggestions	130
Appendices	131
Appendix A. Proof of Theorem 3.2.2	132
Appendix B. Proof of Theorem 4.3.1	134
Bibliography	136
Vita	145

List of Tables

2.1	The model fit as measured by R^2 in (2.55) of identified models for simulation and prediction of validation data for the evaporator. (R^2 less than zero is indicated by '–'.)	34
4.1	The process variables for the boiler process	97
4.2	Fault identification results for the single biased sensor with I_{FSR} through OSR (O) and SRAMS (S) approaches. \checkmark means the faulty sensor can be correctly identified, while $-$ means the faulty sensor cannot be identified.	99
4.3	The minimum bias fault required by fault isolability condition with $\alpha = 0.95$ and $\beta = 0.90$, which is calculated as the percentage of the mean of the corresponding sensor.	99
4.4	Fault identification results for the single precision degraded sensor with $I_{V_{sum}}$ through OSR(O) and SRAMS (S) approaches. \checkmark means the faulty sensor can be correctly identified, while $-$ means the faulty sensor cannot be identified.	100
4.5	The minimum precision degradation fault required by fault isolability condition with $\alpha = 0.95$ and $\beta = 0.90$, which is calculated as the percentage of the standard deviation of the corresponding sensor.	101
4.6	The mean square error of reconstructed measurements with maximum likelihood (ML) and least squares (LS), respectively.	102
5.1	Basic statistics for the outputs of the 4×4 process	121
5.2	Fault identification results for the single biased sensor with I_{FSR} through OSR approach. \checkmark means the faulty sensor can be correctly identified, while $-$ means the faulty sensor cannot be identified.	122
5.3	Fault identification results for the single precision degraded sensor with V_{sum} through OSR approach. \checkmark means the faulty sensor can be correctly identified, while $-$ means the faulty sensor cannot be identified.	123

List of Figures

2.1	Asymptotic pole estimation results of the SISO counter example	28
2.2	Asymptotic zero estimation results of the SISO counter example	29
2.3	The value r defined by (2.54) for the 100 randomly chosen systems as defined in the text. The number of realizations for each system, M , was 25. (The value for system 31 is 4.46 and out of range.) PARSIM-P is better than N4SID/CVA in 84 of the 100 cases. The average excess of standard deviation for N4SID/CVA is 9.1%.	31
3.1	Pole estimates for the SISO simulation example	62
3.2	The estimates of the frequency response from PEM for SISO closed-loop simulations	63
3.3	The estimates of the frequency response from PARSIM-E SISO closed-loop simulations	64
3.4	The estimates of the frequency response from PARSIM-E1 SISO closed-loop simulations	65
3.5	The estimates of the frequency response from N4SID for SISO closed-loop simulations	66
3.6	The N4SID pole estimation for 10 Monte-Carlo closed-loop simulations: \times estimated pole, $+$ system pole	67
3.7	The PARSIM-E pole estimation for 10 Monte-Carlo closed-loop simulations: \times estimated pole, $+$ system pole	68
3.8	The PARSIM-E1 pole estimation for 10 Monte-Carlo closed-loop simulations: \times estimated pole, $+$ system pole	69
3.9	The estimates of the frequency response from N4SID for MIMO closed-loop simulations	70
3.10	The estimates of the frequency response from PARSIM-E for MIMO closed-loop simulations	71
3.11	The estimates of the frequency response from PARSIM-E1 for MIMO closed-loop simulations	72
3.12	The Bode magnitude plot of PARSIM-E for SISO closed-loop simulations.	73

3.13	The eigenvalues of estimated A matrix from PARSIM-E: \times estimated pole, $+$ system pole.	74
3.14	The Bode magnitude plot of PARSIM-E1 for SISO closed-loop simulations.	75
3.15	The eigenvalues of estimated A matrix from PARSIM-E1: \times estimated pole, $+$ system pole.	76
3.16	The Bode magnitude plot of Jansson's approach for SISO closed-loop simulations.	77
3.17	The eigenvalues of estimated A matrix from SSARX: \times estimated pole, $+$ system pole.	78
3.18	The Bode magnitude plot of "whitening filter" approach for SISO closed-loop simulations.	79
3.19	The eigenvalues of estimated A matrix from "whitening filter" approach: \times estimated pole, $+$ system pole.	80
4.1	w_i^j ($j = 2, 3, 4$) is the projection of corresponding b_j ($j = 2, 3, 4$) to S_{w_1} , while w_1 may be orthogonal or nearly orthogonal to b_4	104
5.1	The output measurement for bias fault at the 4 th sensor with fault magnitude as 40% of the standard deviation of the corresponding measurements	124
5.2	The average I_{FSR} and $I_{V_{sum}}$ for bias fault at the 4 th sensor with fault magnitude as 40% of the standard deviation of the corresponding measurements	125
5.3	The output measurement for precision degradation fault at the 1 th sensor with fault magnitude as 25.0% of the standard deviation of the corresponding measurements	126
5.4	The average I_{FSR} and $I_{V_{sum}}$ for precision degradation fault at the 1 th sensor with fault magnitude as 25.0% percentage of the standard deviation of the corresponding measurements	127

Chapter 1

Introduction and Dissertation Outline

The development of system identification and fault diagnosis theory is of great practical significance. Systems are concerned with a broad spectrum of human-made machinery, including industrial production facilities (power plants, chemical plants, oil refinery, semiconductor fabrication plants, steel mills, paper mills, etc.), transportation vehicles (ships, airplanes, automobiles) and household appliances (heating/air conditioning equipment, refrigerators, washing machines, etc.).

System identification deals with the problem of building systems models based on observed data from the system. Two landmark papers [3, 26] gave birth to the Prediction Error Identification framework and the Subspace Identification framework, respectively. The advantage of prediction error methods (PEMs) is that the convergence and asymptotic variance results are available [43], which are important for "identification for control" applications [25]. The disadvantage of PEMs is that a complicated parametrization step is involved for MIMO systems, which makes them difficult to apply in practice.

The motivation of circumventing the complicated parametrization of PEMs, especially for the MIMO identification, leads to tremendous interest

in subspace identification methods (SIMs). Most SIMs fall into the unifying theorem proposed by Van Overschee and De Moor [64], among which are canonical variate analysis (CVA) [37], N4SID [63], subspace fitting [30], and MOESP [68]. Based on the unifying theorem, all these algorithms can be interpreted as a singular value decomposition of a weighted matrix.

SIMs have many advantages as an alternative to the more traditional prediction error method or maximum likelihood (ML) approach and they are very good for delivering initial estimates to PEM. A few drawbacks have been experienced with SIMs:

1. The estimation accuracy in general is not as good as the PEM in terms of the variance of the estimates.
2. The application of SIMs to closed-loop data typically gives biased estimates, even though the data satisfy identifiability conditions for traditional methods such as PEMs.

In this work, we are concerned with the reasons why subspace identification approaches exhibit these drawbacks and propose new SIMs which use fewer estimated parameters (i.e., more parsimonious).

On the other hand, fault diagnosis involves early detection and isolation of faults, which is critical in avoiding product quality deterioration, performance degradation, major damage to the equipment and hazard to human health or even loss of lives. The traditional approaches to fault detection

and diagnosis involve limit checking of some key variables or the application of redundant sensors (physical redundancy). Over the last two decades, fault detection and diagnosis have gained increasing consideration world-wide. This development was mainly stimulated by the trend of automation towards more complexity and the growing demand for higher security of control systems.

Advanced methods can be divided into two categories as qualitative model (knowledge model) based approach and quantitative model based approach. The objective of qualitative model based approach is to identify the symptoms corresponding to the observations of the process that can be used for a fault decision on the basis of the knowledge redundancy, such as neural networks, expert systems and fuzzy logic. In the field of the quantitative model based approach, a strong impetus comes from the side of modern control theory that has brought forth mathematical modeling, state estimation and parameter identification that have been made feasible by the progress of modern computer technology.

The model based approach can be related to chemical process engineering, where the traditional material and energy balance calculations evolved into systematic data reconciliation and the detection of gross errors. The work in this area is reviewed thoroughly by Crowe [13]. Another root can be traced to aerospace related research, which leads to the fundamental formulation of parity relation concepts [12, 47]. An important related activity is due to Gertler and coworkers [20, 21], who try to diagnose faults by designing structured residuals that are insensitive to a particular subset of faults. In par-

allel, and partially overlapping with the above efforts, several researchers were looking into the possibility of applying Kalman filters [33, 41] and diagnostic observers [18, 19, 49] to fault detection and isolation problem. In the area of fault detection and isolation by parameter estimation, substantial work has been done by Isermann and colleagues [28]. An important related activity is due to Basseville and coworkers, concerning the detection of small parametric faults by the statistical analysis of residuals obtained over extended sets of observations [4, 5].

Motivated by early work by Gertler [20] and Qin and Li [55], we propose a new optimal structured residuals approach for improved fault diagnosis. To maximize fault isolation ability, a matrix of optimal structured residuals are designed. Each of them is insensitive to one subset of faults while being most sensitive to one of remaining ones. The maximum of all structured residuals in each row is then selected as the optimal one for fault isolation. Through this approach, optimal structured residual directions with maximum fault isolation ability are obtained.

The dissertation is organized as follows:

In Chapter 2, a novel subspace identification approach is proposed to enforce the causality of high order ARX models. The key idea is to avoid the estimation of parameters that are known to be zero. This means that a lower triangular structure of an estimated matrix must be enforced which leads to somewhat more complicated calculations. The new algorithms, which fall into the subspace fitting framework, are shown to be consistent under mild

assumptions and applicable to a general state space model structure.

Chapter 3 investigates the reason why traditional SIMs can not handle data under closed-loop condition. In the chapter, we show that the closed-loop consistency with SIMs can be achieved through innovation estimation. Based on this analysis, a new SIM with parsimonious formulation is proposed to handle data collected under feedback.

In Chapter 4, the possibility of misidentification of faults with traditional structured residuals approaches is investigated. Based on this analysis a new optimal structured residuals (OSR) design criterion for unidirectional fault isolation is proposed. To maximize fault isolation ability, a matrix of optimal structured directions are designed. Each of them is insensitive to one particular fault while being most sensitive to one of the remaining ones.

Chapter 5 extends the optimal structured residuals (OSR) approach to multidimensional fault cases. Faults occurred in dynamic systems can be considered as well using the extended state space model or the dynamic principal component model.

Chapter 6 summarizes the present research results and provides future research possibilities in these areas.

Chapter 2

A Novel Subspace Identification Approach with Enforced Causal Models

In this chapter, we are concerned with the reasons why subspace identification approaches exhibit large estimation variance and propose new SIMs which use fewer estimated parameters (i.e., more parsimonious) for open loop applications. First of all, we start with the analysis of existing subspace formulation using the linear regression formulation [31, 34]. This means that essentially several ARX models are estimated directly from data with different prediction intervals. From this analysis we reveal that the typical SIM algorithms use extra terms in the model that appear to be non-causal. These terms, although conveniently included for performing subspace projections, are the causes for inflated variance in the estimates and partially responsible for the loss of closed-loop identifiability. Peternell et al. [51] observe this point as well and use constrained least squares to improve the estimate.

The rest of the chapter is organized as follows. In Section 2.1, we analyze the existing SIMs and point out the non-causal projection. Based on this observation, novel SIM formulations with only causal terms are presented in detail in Section 2.2. Numerical implementation of proposed algorithms is

introduced in Section 2.3. In Section 2.4, numerical simulations are given to show the efficiency of the proposed algorithm. Section 3.5 summarizes the chapter.

2.1 Analysis of subspace formulation

2.1.1 Problem formulation and assumptions

We assume that the system to be identified can be written in an innovation form as

$$x_{k+1} = Ax_k + Bu_k + Ke_k \quad (2.1a)$$

$$y_k = Cx_k + Du_k + e_k \quad (2.1b)$$

where $y_k \in R^{n_y}$, $x_k \in R^n$, $u_k \in R^{n_u}$, and $e_k \in R^{n_y}$ are the system output, state, input, and innovation, respectively. A , B , C and D are system matrices with appropriate dimensions. K is the Kalman filter gain. To establish statistical consistency of the SIM, we introduce following assumptions:

A1 : The eigenvalues of $A - KC$ are strictly inside the unit circle.

A2 : The system is minimal in the sense that (A, C) is observable and $(A, [B, K])$ is controllable.

A3 : The innovation sequence e_k is a stationary, zero mean, white noise process with ergodic second order moments

$$E(e_i e_j^T) = R \delta_{ij}$$

where δ_{ij} is the Kronecker delta.

A4 : The input u_k and innovation sequence e_j are uncorrelated for $\forall k$ and $\forall j$, i.e., the system operates in open loop.

A5 : The input signal is quasi-stationary [45] and is persistently exciting of order $f+p$, where f and p stand for future and past horizons, respectively, to be defined later.

The identification problem is: given a set of input/output measurements, estimate the system matrices (A, B, C, D) , Kalman filter gain K up to within a similarity transformation and the innovation covariance matrix R .

Based on the state space description in (2.1), an extended state space model can be formulated as

$$Y_f = \Gamma_f X_k + H_f U_f + G_f E_f \quad (2.2a)$$

$$Y_p = \Gamma_p X_{k-p} + H_p U_p + G_p E_p \quad (2.2b)$$

where the subscripts f and p denote future and past horizons, respectively.

The extended observability matrix is

$$\Gamma_f = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{f-1} \end{bmatrix} \quad (2.3)$$

and H_f and G_f are Toeplitz matrices:

$$H_f = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-2}B & CA^{f-3}B & \cdots & D \end{bmatrix} \quad (2.4a)$$

$$G_f = \begin{bmatrix} I & 0 & \cdots & 0 \\ CK & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-2}K & CA^{f-3}K & \cdots & I \end{bmatrix} \quad (2.4b)$$

The input and output data are arranged in the following Hankel form:

$$U_f = \begin{bmatrix} u_k & u_{k+1} & \cdots & u_{k+N-1} \\ u_{k+1} & u_{k+2} & \cdots & u_{k+N} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k+f-1} & u_{k+f} & \cdots & u_{k+f+N-2} \end{bmatrix} \quad (2.5a)$$

$$\triangleq \begin{bmatrix} u_f(k) & u_f(k+1) & \cdots & u_f(k+N-1) \end{bmatrix} \quad (2.5b)$$

$$U_p = \begin{bmatrix} u_{k-p} & u_{k-p+1} & \cdots & u_{k-p+N-1} \\ u_{k-p+1} & u_{k-p+2} & \cdots & u_{k-p+N} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k-1} & u_k & \cdots & u_{k+N-2} \end{bmatrix} \quad (2.5c)$$

$$\triangleq \begin{bmatrix} u_p(k-p) & u_p(k-p+1) & \cdots & u_p(k-p+N-1) \end{bmatrix} \quad (2.5d)$$

Similar formulations are made for Y_f , Y_p , E_f , and E_p . The state sequences are defined as:

$$X_k = [x_k, x_{k+1}, \cdots, x_{k+N-1}] \quad (2.6a)$$

$$X_{k-p} = [x_{k-p}, x_{k-p+1}, \cdots, x_{k-p+N-1}] \quad (2.6b)$$

Subspace identification consists of estimating the extended observability matrix first and then the model parameters.

2.1.2 Analysis of conventional SIMs

As the first step, subspace identification methods minimize the following objective function [65],

$$\begin{aligned} [\hat{L}^1 \quad \hat{L}^2 \quad \hat{L}^3] &= \arg \min \{ \|Y_f - L^1 Y_p - L^2 U_p - L^3 U_f\|_F^2 \} \\ &= \arg \min \left\{ \sum_{j=0}^{N-1} \left\| y_f(k+j) - [L^1 \quad L^2 \quad L^3] \begin{bmatrix} y_p(k-p+j) \\ u_p(k-p+j) \\ u_f(k+j) \end{bmatrix} \right\|^2 \right\} \end{aligned} \quad (2.7)$$

where u_f , u_p , y_f , and y_p are defined in (2.5b) and (2.5d) as columns of the corresponding data matrices.

Denoting

$$L^1 = \begin{bmatrix} L_{11}^1 & L_{12}^1 & \cdots & L_{1p}^1 \\ L_{21}^1 & L_{22}^1 & \cdots & L_{2p}^1 \\ \vdots & & \ddots & \\ L_{f1}^1 & L_{f1}^1 & & L_{fp}^1 \end{bmatrix} \triangleq \begin{bmatrix} L_1^1 \\ L_2^1 \\ \vdots \\ L_f^1 \end{bmatrix} \quad (2.8a)$$

$$L^2 = \begin{bmatrix} L_{11}^2 & L_{12}^2 & \cdots & L_{1p}^2 \\ L_{21}^2 & L_{22}^2 & \cdots & L_{2p}^2 \\ \vdots & & \ddots & \\ L_{f1}^2 & L_{f1}^2 & & L_{fp}^2 \end{bmatrix} \triangleq \begin{bmatrix} L_1^2 \\ L_2^2 \\ \vdots \\ L_f^2 \end{bmatrix} \quad (2.8b)$$

$$L^3 = \begin{bmatrix} L_{11}^3 & L_{12}^3 & \cdots & L_{1f}^3 \\ L_{21}^3 & L_{22}^3 & \cdots & L_{2f}^3 \\ \vdots & & \ddots & \\ L_{f1}^3 & L_{f1}^3 & & L_{ff}^3 \end{bmatrix} \triangleq \begin{bmatrix} L_1^3 \\ L_2^3 \\ \vdots \\ L_f^3 \end{bmatrix} \quad (2.8c)$$

the above problem is equivalent to f separate sub-problems:

$$\begin{aligned} [\hat{L}_i^1 \quad \hat{L}_i^2 \quad \hat{L}_i^3] &= \\ \arg \min \left\{ \sum_{j=0}^{N-1} \left\| y_{k+j+i-1} - [L_i^1 \quad L_i^2 \quad L_i^3] \begin{bmatrix} y_p(k-p+j) \\ u_p(k-p+j) \\ u_f(k+j) \end{bmatrix} \right\|^2 \right\} \end{aligned} \quad (2.9)$$

for $i = 1, \dots, f$, this is to say that f different ARX models are estimated from data. Consider the i^{th} subproblem and spell out the nature of the term

$L_i^3 u_f(k+j)$. This subproblem corresponds to the model

$$\begin{aligned}
y_{k+i-1} &= \begin{bmatrix} L_i^1 & L_i^2 \end{bmatrix} \begin{bmatrix} y_p(k-p) \\ u_p(k-p) \end{bmatrix} + L_i^3 u_f(k) + v_k \\
&= \begin{bmatrix} L_i^1 & L_i^2 \end{bmatrix} \begin{bmatrix} y_p(k-p) \\ u_p(k-p) \end{bmatrix} + L_{i1}^3 u_k + L_{i2}^3 u_{k+1} + \dots L_{ii}^3 u_{k+i-1} \\
&\quad + \sum_{j=i+1}^f L_{ij}^3 u_{k+j-1} + v_k
\end{aligned} \tag{2.10}$$

Note that the summation in (2.10) represents a non-causal relation from u to y . That is, L_{ij}^3 are estimated even though it is known that $L_{ij}^3 = 0$ for $j > i$. The matrix L^3 is, in other words, *block lower triangular*. However, this information is not normally taken care of in (2.7), as pointed out in [61]. While there is no problem from a consistency point of view given proper excitation of the input, known parameters are estimated from data. Therefore, we can make the following statements about the typical SIM formulation in general.

1. The model format used in SIM during the projection step is non-causal. This would result in non-causal models in the projection step. Although the non-causal terms are ignored at the step to estimate B , D , all the model parameters estimate have inflated variance due to the fact that extra and unnecessary terms are included in the model.
2. Because of the extra terms that turn out to be ‘future’ inputs relative to the output, SIMs in general have problems with closed-loop data using direct identification methods. Most SIMs usually project out U_f as follows:

$$Y_f \Pi_{U_f}^\perp = \Gamma_f X_k \Pi_{U_f}^\perp + G_f E_f \Pi_{U_f}^\perp \tag{2.11}$$

where $\Pi_{U_f}^\perp = I - U_f^T(U_f U_f^T)^{-1}U_f$. Because of the non-causal terms in the model, $\frac{1}{N}E_f U_f^T \neq 0$ as $N \rightarrow \infty$ for closed-loop data. As a consequence, many SIMs fail to work on closed loop data, except for a few SIM algorithms that avoid this projection [11, 69].

3. Because U_f contains extra rows due to the extra terms, the projection in (2.11) tends to reduce the information content unnecessarily even for open-loop data, leading to inefficient use of the data.
4. These non-causal terms will have negligible coefficients only when the number of data is very large and process is well excited. For a limited number of samples or non-white input signals, SIM algorithms tend to have large estimation errors.

To avoid these problems the SIM model must not include these non-causal terms, Peternell et al. [51] propose a few methods to exclude these extra terms. Specially, they recommend a two steps procedure; (i) use a conventional (unconstrained) SIM to estimate the deterministic Markov parameters $CA^{i-1}B$; and (ii) form H_f with these Markov parameters to ensure that it is lower triangular and then estimate the extended observability matrix. We propose a parallel and a sequential implementation of a causal subspace identification method (PARSIM) which remove these non-causal terms by enforcing a lower triangular structure in L^3 and hence of H_f at every step of the SIM procedure. By enforcing a lower-triangular structure, we reduce the number of estimated parameters in this stage by $f(f-1)/2$. The parallel

PARSIM (PARSIM-P) method involves a bank of least squares (LS) problems in parallel, while the sequential PARSIM (PARSIM-S) involves a bank of LS problems sequentially. Optimal weighting is derived for the PARSIM algorithms. An optimal estimate of the B, D matrices is given using the Kalman filter structure.

2.2 Subspace identification avoiding non-causal terms

The key idea in the proposed method is to exclude the non-causal terms of U_f mentioned in Section 2.1. To accomplish this we partition the extended state space model row-wise as follows:

$$Y_f = \begin{bmatrix} Y_{f1} \\ Y_{f2} \\ \vdots \\ Y_{ff} \end{bmatrix}; \quad Y_i \triangleq \begin{bmatrix} Y_{f1} \\ Y_{f2} \\ \vdots \\ Y_{fi} \end{bmatrix}; i = 1, 2, \dots, f \quad (2.12)$$

where $Y_{fi} = [y_{k+i-1} \quad y_{k+i} \quad \dots \quad y_{k+N+i-2}]$. Partition U_f and E_f in a similar way to define U_{fi} , U_i , E_{fi} , and E_i , respectively, for $i = 1, 2, \dots, f$. Denote further

$$\Gamma_f = \begin{bmatrix} \Gamma_{f1} \\ \Gamma_{f2} \\ \vdots \\ \Gamma_{ff} \end{bmatrix} \quad (2.13a)$$

$$H_{fi} \triangleq [CA^{i-2}B \quad \dots \quad CB \quad D] \quad (2.13b)$$

$$\triangleq [H_{i-1} \quad \dots \quad H_1 \quad H_0] \quad (2.13c)$$

$$G_{fi} \triangleq [CA^{i-2}K \quad \dots \quad CK \quad I] \quad (2.13d)$$

$$\triangleq [G_{i-1} \quad \dots \quad G_1 \quad G_0] \quad (2.13e)$$

where $\Gamma_{fi} = CA^{i-1}$, and H_i and G_i are the Markov parameters for the deterministic input and innovation sequence, respectively. We have the following equations by partitioning (2.2a),

$$Y_{fi} = \Gamma_{fi}X_k + H_{fi}U_i + G_{fi}E_i \quad (2.14)$$

for $i = 1, 2, \dots, f$. Note that each of the above equations is guaranteed causal.

2.2.1 PARSIM algorithms

By eliminating $e(k)$ in the innovation model through iteration, it is straightforward to derive the following relation [34],

$$X_k = L_z Z_p + A_K^p X_{k-p} \quad (2.15)$$

where

$$L_z \triangleq \begin{bmatrix} \Delta_p(A_K, K) & \Delta_p(A_K, B_K) \end{bmatrix} \quad (2.16a)$$

$$\Delta_p(A, B) \triangleq \begin{bmatrix} A^{p-1}B & \dots & AB & B \end{bmatrix} \quad (2.16b)$$

$$A_K \triangleq A - KC \quad (2.16c)$$

$$B_K \triangleq B - KD \quad (2.16d)$$

$$Z_p \triangleq \begin{bmatrix} Y_p^T & U_p^T \end{bmatrix}^T \quad (2.16e)$$

Substituting this equation into (2.14), we obtain

$$Y_{fi} = \Gamma_{fi}L_z Z_p + \Gamma_{fi}A_K^p X_{k-p} + H_{fi}U_i + G_{fi}E_i \quad (2.17)$$

for $i = 1, 2, \dots, f$. Note the second term on the RHS of (2.17) tends to zero as p tends to infinity under Assumption A1. Now we have the following parallel PARSIM algorithm to estimate Γ_{fi} and H_{fi} .

[Algorithm 1] Parallel PARSIM (PARSIM-P)

1. Perform the following LS estimates, for $i = 1, 2, \dots, f$,

$$\begin{bmatrix} \hat{\Gamma}_{fi}L_z & \hat{H}_{fi} \end{bmatrix} = Y_{fi} \begin{bmatrix} Z_p \\ U_i \end{bmatrix}^\dagger \quad (2.18)$$

where $[\cdot]^\dagger$ stands for the Moore-Penrose pseudo-inversion. Stack $\hat{\Gamma}_{fi}L_z$ together to obtain $\hat{\Gamma}_fL_z$ as

$$\begin{bmatrix} \hat{\Gamma}_{f1}L_z \\ \hat{\Gamma}_{f2}L_z \\ \vdots \\ \hat{\Gamma}_{ff}L_z \end{bmatrix} = \hat{\Gamma}_fL_z \quad (2.19)$$

2. Perform SVD for the following weighted matrix

$$W_1(\hat{\Gamma}_fL_z)W_2 = U_nS_nV_n^T + \varepsilon \quad (2.20)$$

where W_1 is nonsingular and L_zW_2 does not lose rank. U_n , S_n and V_n are associated to the first n largest singular value. The residual term ε stands for the product of the remaining singular vectors and singular values. We choose

$$\hat{\Gamma}_f = W_1^{-1}U_nS_n^{1/2} \quad (2.21)$$

from which the estimate of A and C can be obtained [67].

3. The estimate of B and D is discussed in the next section using a Kalman filter formulation.

Notice that the proposed parallel PARSIM gives *consistent* estimates for Γ_f and $H_{i-1}, \forall i = 1, 2, \dots, f$ under the assumptions stated in Section 2.1. To rationalize the statement, it is sufficient to show that as $N \rightarrow \infty$,

$$\begin{bmatrix} \hat{\Gamma}_{fi} L_z & \hat{H}_{fi} \end{bmatrix} \rightarrow \begin{bmatrix} \Gamma_{fi} L_z & H_{fi} \end{bmatrix}$$

where $\begin{bmatrix} \hat{\Gamma}_{fi} L_z & \hat{H}_{fi} \end{bmatrix}$ is calculated according to (2.18). Assumption A1 implies that the initial state has negligible effect on the estimate with sufficient large p , as shown in (2.17). From A4 we have $\frac{1}{N} E_i Z_p^T \rightarrow 0$ and $\frac{1}{N} E_i U_i^T \rightarrow 0$ as $N \rightarrow \infty$. Substituting (2.17) with $p \rightarrow \infty$ into (2.18) leads to

$$\begin{aligned} \begin{bmatrix} \hat{\Gamma}_{fi} L_z & \hat{H}_{fi} \end{bmatrix} &= \begin{bmatrix} \Gamma_{fi} L_z & H_{fi} \end{bmatrix} \begin{bmatrix} Z_p \\ U_i \end{bmatrix} \begin{bmatrix} Z_p \\ U_i \end{bmatrix}^\dagger + G_{fi} E_i \begin{bmatrix} Z_p \\ U_i \end{bmatrix}^\dagger \\ &= \begin{bmatrix} \Gamma_{fi} L_z & H_{fi} \end{bmatrix} + G_{fi} \left(\frac{1}{N} E_i \begin{bmatrix} Z_p \\ U_i \end{bmatrix}^T \right) \left(\frac{1}{N} \begin{bmatrix} Z_p \\ U_i \end{bmatrix} \begin{bmatrix} Z_p \\ U_i \end{bmatrix}^T \right)^{-1} \\ &\rightarrow \begin{bmatrix} \Gamma_{fi} L_z & H_{fi} \end{bmatrix} \end{aligned}$$

as $N \rightarrow \infty$. Assumption A5 guarantees that all system modes are sufficiently excited so that the matrix inverse in the above equation exists. It has been shown in [34] that A2 is needed for L_z to have full row rank and Γ_f to have full column rank. Therefore, the SVD step in the PARSIM-P algorithm guarantee that $\hat{\Gamma}_f$ and Γ_f have the same column space asymptotically.

The PARSIM-P algorithm estimates the model parameters in parallel which re-estimate some of the Markov parameters in H_{fi} repeatedly. To avoid

this we rewrite (2.17) by ignoring the A_K^p term

$$Y_{fi} = \Gamma_{fi} L_z Z_p + H_{i-1} U_{f1} + H_{f(i-1)} [U_{f2}^T \cdots U_{fi}^T]^T + G_{fi} E_i$$

where H_{i-1} is defined in (2.13c). If we perform the above projections sequentially for $i = 1, 2, \dots, f$, $H_{f(i-1)}$ is estimated in the $(i-1)^{th}$ step. Γ_{fi} and H_{i-1} are the only unknown at the i^{th} step.

[Algorithm 2] Sequential PARSIM (PARSIM-S)

1. Perform the following LS for $i = 1$,

$$\begin{bmatrix} \hat{\Gamma}_{f1} L_z & \hat{H}_{f1} \end{bmatrix} = Y_{f1} \begin{bmatrix} Z_p \\ U_{f1} \end{bmatrix}^\dagger \quad (2.23)$$

2. Perform the following causal projection for $i = 2, \dots, f$

$$\begin{bmatrix} \hat{\Gamma}_{fi} & \hat{H}_{i-1} \end{bmatrix} = (Y_{fi} - \hat{H}_{f(i-1)}) [U_{f2}^T \cdots U_{fi}^T]^T \begin{bmatrix} Z_p \\ U_{f1} \end{bmatrix}^\dagger \quad (2.24)$$

Stack $\hat{\Gamma}_{fi} L_z$ together as equation (2.19).

3. Same as the Step 2 in Algorithm 1.

The sequential PARSIM gives *consistent* estimates for Γ_f and H_{i-1} , $\forall i = 1, 2, \dots, f$ under the assumptions stated in Section 2.1. The proof is similar to that of PARSIM-P, therefore we omit it in the paper.

[Remark 1] For finite past horizon p the algorithm is biased, but the bias decays to zero exponentially with p . If p is too large in practice, however, large variance is expected for the estimates. Therefore, it is necessary in

practice to use a finite p for the best trade-off. Cross-validation can be used to select an optimal p .

[Remark 2] The parallel PARSIM requires that no correlation exists between future u_k and past e_k to be consistent, which is only valid under open loop condition, therefore the PARSIM-P algorithms are biased for direct closed loop identification. To make it applicable to closed-loop data, an innovation estimation approach is proposed in [58].

[Remark 3] The Markov parameters, $H_{i-1}, \forall i = 1, 2, \dots, f$, can be estimated directly from the SIMs without the knowledge of system matrices, (A, B, C, D) . Meanwhile the low triangular structure of the Toeplitz matrix, H_f , is conserved.

2.2.2 Improved variance of PARSIM algorithms

After presenting the PARSIM algorithms, we analyze the variance of the PARSIM estimates relative to that of conventional SIM algorithms. For conventional SIMs the asymptotic variance of the model estimates is derived in [6–8]. These analyses provide insight into what contribute to the variance of the estimates.

In this subsection we provide a covariance equality for PARSIM estimates by interpreting the subspace projections in the generalized least squares (GLS) framework [48]. For the i^{th} block-row we explained that conventional SIMs use model (2.10) but the process is actually (2.17). By comparing (2.10)

with (2.17) when p is large we have,

$$\begin{aligned} \begin{bmatrix} L_i^1 & L_i^2 \end{bmatrix} &= \Gamma_{fi} L_z \\ L_i^3 &= \begin{bmatrix} H_{fi} & \vdots & 0 & \cdots & 0 \end{bmatrix} \\ v_k &= \sum_{j=1}^i G_{i-j} e_{k+j-1} \end{aligned}$$

Note that v_k is auto-correlated, therefore the SIM projections do not fit into the maximum likelihood framework. Denoting

$$\begin{aligned} cov[v_k] &= \Sigma_v \\ V_{fi} &= G_{fi} E_i \end{aligned}$$

and

$$\mathbb{V}_{i,N} = vec(V_{fi}^T)$$

where $vec()$ of a matrix forms a long column vector by stacking the columns of that matrix, we have

$$cov(\mathbb{V}_{i,N}) = \Sigma_v \otimes I_N$$

With this notation we can convert the PARSIM equation (2.17) into

$$\mathbb{Y}_{i,N} = \mathbb{X}_{1,N} \theta_1 + \mathbb{V}_{i,N} \tag{2.25}$$

where

$$\begin{aligned} \mathbb{Y}_{i,N} &= vec(Y_{fi}^T) \\ \mathbb{X}_{1,N} &= I \otimes \begin{bmatrix} Z_p^T & U_i^T \end{bmatrix} \\ \theta_1 &= vec(\begin{bmatrix} \Gamma_{fi} L_z & H_{fi} \end{bmatrix}^T) \end{aligned}$$

Similarly, the conventional SIM equation for the i^{th} block row (2.10) can be converted to

$$\mathbb{Y}_{i,N} = \mathbb{X}_{1,N}\theta_1 + \mathbb{X}_{2,N}\theta_2 + \mathbb{V}_{i,N} \quad (2.26)$$

where

$$\mathbb{X}_{2,N} = I \otimes [U_{f(i+1)}^T \quad U_{f(i+2)}^T \quad \cdots \quad U_{ff}^T]$$

is the matrix of non-causal input data and

$$\theta_2 = vec([L_{i(i+1)}^3 \quad L_{i(i+2)}^3 \cdots L_{if}^3]^T)$$

is the vector of extra parameters in conventional SIMs. Now we state that the least squares solutions (2.17) for PARSIMs and (2.10) for conventional SIMs are identical to the GLS solution to (2.25) and (2.26), respectively [48]. The estimates from both conventional SIMs and PARSIMs are consistent, which is not concerned here. The question is whether PARSIM estimates have smaller variance than conventional SIMs regardless of the data length N .

From [48] we know that the GLS interpretation of PARSIM estimates leads to

$$cov(\hat{\theta}_{1,N}) = (\mathbb{X}_{1,N}^T (\Sigma_v \otimes I_N)^{-1} \mathbb{X}_{1,N})^{-1}$$

where $\hat{\theta}_{1,N}$ is the PARSIM estimate for θ_1 and

$$cov \left(\begin{bmatrix} \hat{\theta}'_{1,N} \\ \hat{\theta}'_{2,N} \end{bmatrix} \right) = \left([\mathbb{X}_{1,N} \quad \mathbb{X}_{2,N}]^T (\Sigma_v \otimes I_N)^{-1} [\mathbb{X}_{1,N} \quad \mathbb{X}_{2,N}] \right)^{-1}$$

where $\hat{\theta}'_{1,N}$ is the conventional SIM estimate for θ_1 .

To simplify the notation, we denote

$$S_{ij,N} = \mathbb{X}_{i,N}(\Sigma_v \otimes I_N)^{-1}\mathbb{X}_{j,N} \quad (2.27)$$

for $i, j = 1, 2$. Then the covariance expressions become

$$\begin{aligned} \text{cov}(\hat{\theta}_{1,N}) &= S_{11,N}^{-1} \\ \text{cov} \left(\begin{bmatrix} \hat{\theta}'_{1,N} \\ \hat{\theta}'_{2,N} \end{bmatrix} \right) &= \begin{bmatrix} S_{11,N} & S_{12,N} \\ S_{12,N}^T & S_{22,N} \end{bmatrix}^{-1} = \begin{bmatrix} \Phi_{11,N} & \Phi_{12,N} \\ \Phi_{12,N}^T & \Phi_{22,N} \end{bmatrix} \end{aligned}$$

from which it is easy to show that

$$\text{cov}(\hat{\theta}'_{1,N}) = \Phi_{11,N} = S_{11,N}^{-1} + S_{11,N}^{-1}S_{12,N}\Phi_{22,N}S_{12,N}^TS_{11,N}^{-1}$$

Therefore

$$\text{cov}(\hat{\theta}'_{1,N}) - \text{cov}(\hat{\theta}_{1,N}) = S_{11,N}^{-1}S_{12,N}\Phi_{22,N}S_{12,N}^TS_{11,N}^{-1} \quad (2.28)$$

Noticing that $\Phi_{22,N}$ is strictly positive definite due to the inverse of the covariance matrix, we have

$$\text{cov}(\hat{\theta}'_{1,N}) \geq \text{cov}(\hat{\theta}_{1,N}) \quad (2.29)$$

regardless of N and the equality holds only if $S_{12,N} = 0$. It is noted further that $S_{11,N}^{-1}S_{12,N}$ in (2.28) is the regression coefficient matrix of $\mathbb{X}_{2,N}$ on $\mathbb{X}_{1,N}$, which is not zero for colored inputs. We can only compare the variance of $\hat{\theta}_1$ rigorously as shown above. For $\hat{\Gamma}_f$ estimate from $\hat{\theta}_1$ we can only say that PARSIM estimate likely leads to a better estimate of the true observability subspace, but we cannot compare the variance since it depends on the basis. Similarly we cannot compare the variance of the system matrices such as C and A .

A reduced variance in estimating the observability matrix will likely lead to better estimates for A and C later. The Monte-Carlo study in Section 2.4 provides strong indications that this indeed is the case.

2.2.3 Determination of observability matrix

In the conventional SIM formulation under open-loop conditions,

$$E_f \Pi_{U_f}^\perp \rightarrow E_f \text{ as } N \rightarrow \infty \quad (2.30)$$

since E_f is uncorrelated with U_f . Therefore, for large N (2.11) becomes:

$$Y_f \Pi_{U_f}^\perp \approx \Gamma_f X_k \Pi_{U_f}^\perp + G_f E_f \quad (2.31)$$

Post-multiplying Z_p^T to (2.31) eliminate the noise term for large N ,

$$Y_f \Pi_{U_f}^\perp Z_p^T \approx \Gamma_f X_k \Pi_{U_f}^\perp Z_p^T \quad (2.32)$$

Van Overschee and De Moor [64] show that all SIM methods perform SVD on the following weighted matrix:

$$W_r Y_f \Pi_{U_f}^\perp Z_p^T W_c = W_r \Gamma_f X_k \Pi_{U_f}^\perp Z_p^T W_c \quad (2.33)$$

where W_r and W_c are the row and column weighting matrices, respectively. In CVA $W_r = (Y_f \Pi_{U_f}^\perp Y_f^T)^{-1/2}$ which basically normalizes the output variables. Gustafsson [23] shows that an approximately optimal weighting for W_c is

$$\begin{aligned} W_c &= (Z_p Z_p^T - Z_p U_f^T (U_f U_f^T)^{-1} U_f Z_p^T)^{-1/2} \\ &= (Z_p \Pi_{U_f}^\perp Z_p^T)^{-1/2} \end{aligned} \quad (2.34)$$

which is used in CVA and MOESP. Substituting (2.34) into (2.33), and replacing X_k with $L_z Z_p$ as instrumental variables, we obtain,

$$\begin{aligned} W_r Y_f \Pi_{U_f}^\perp Z_p^T W_c &= W_r \Gamma_f L_z Z_p \Pi_{U_f}^\perp Z_p^T (Z_p \Pi_{U_f}^T Z_p^T)^{-1/2} \\ &= W_r \Gamma_f L_z (Z_p \Pi_{U_f}^\perp Z_p^T)^{1/2} \end{aligned} \quad (2.35)$$

Comparing (2.35) with (2.20), the equivalent weightings for the PARSIMs algorithm are

$$W_1 = W_r \quad (2.36)$$

$$W_2 = (Z_p \Pi_{U_f}^\perp Z_p^T)^{1/2} \quad (2.37)$$

Gustafsson and Rao [24] show that the row-weighting W_1 has no influence on the asymptotic accuracy of the estimated observability matrix. Our simulation experience shows that W_1 has negligible influence on the accuracy of the estimated system matrices as well. Therefore, we suggest to use $W_1 = I$ in the PARSIM algorithms.

2.3 Numerical implementation of PARSIMs

Since the projections in the PARSIM algorithms bear similarity to the standard SIMs such as MOESP, it is straightforward to implement these parallel or sequential projections using QR decomposition [59]. In this section, a new approach to calculate the B , D matrices is derived by prewhitening the equation error of the general state space model.

2.3.1 QR implementation for K

Once $\hat{\Gamma}_f$ is known, the Kalman filter gain K can be estimated [14].

With a large p , substituting (2.15) into (2.2) leads to:

$$Y_f = \Gamma_f L_z Z_p + H_f U_f + G_f E_f \quad (2.38)$$

Therefore,

$$Y_f \Pi_{\begin{smallmatrix} Z_p \\ U_f \end{smallmatrix}}^\perp = G_f E_f \Pi_{\begin{smallmatrix} Z_p \\ U_f \end{smallmatrix}}^\perp = G_f E_f \quad (2.39)$$

since E_f is not correlated with Z_p and U_f in open-loop. Performing QR decomposition,

$$\begin{bmatrix} Z_p \\ U_f \\ Y_f \end{bmatrix} = \begin{bmatrix} R_{11} & & \\ R_{21} & R_{22} & \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} \quad (2.40)$$

then

$$R_{33} Q_3 = G_f E_f \quad (2.41)$$

Denoting $e_k = F e_k^*$ such that $\text{cov}(e_k^*) = I$, from Assumption A3 we have $FF^T = R$. Using this notation we have

$$G_f E_f = G_f^* E_f^* \quad (2.42)$$

where

$$G_f^* = \begin{bmatrix} F & 0 & \cdots & 0 \\ CKF & F & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-2}KF & CA^{f-3}KF & \cdots & F \end{bmatrix} \in \Re^{n_y f \times n_y f} \quad (2.43)$$

From equation (2.41) and (2.42) and using the fact that Q_3 is an orthonormal matrix, we choose

$$\hat{E}_f^* = Q_3 \quad (2.44a)$$

$$\hat{G}_f^* = R_{33} \quad (2.44b)$$

Therefore,

$$\hat{F} = R_{33}(1 : n_y, 1 : n_y) \quad (2.45)$$

and K can be calculated from G_f^* using Γ_f .

2.3.2 Determination of B, D

With A and C estimates, Section 10.6 in [45] gives an effective approach to estimate B and D with an output error formulation. Note that there is a choice whether or not to prewhiten the residuals, as discussed, e.g., in [44]. This choice also corresponds to whether '**focus**' is set to '**simulation**' or '**prediction**' (default) in the N4SID function of the System Identification Toolbox. Here we give a modified approach to estimating B, D and the initial state optimally using A, C, K and F for the general innovation form. Since the initial state is estimated this step does not introduce a bias for finite p .

From the innovation form of the system we have:

$$x_{k+1} = A_K x_k + B_K u_k + K y_k \quad (2.46)$$

where A_k and B_k are defined in (2.16). The process output can be represented

as

$$\begin{aligned} y_k &= C(qI - A_K)^{-1}x_0 + [C(qI - A_K)^{-1}B_K + D]u_k \\ &\quad + C(qI - A_K)^{-1}Ky_k + e_k \end{aligned} \quad (2.47)$$

or:

$$\begin{aligned} [I - C(qI - A_K)^{-1}K]y_k &= C(qI - A_K)^{-1}x_0 \\ &\quad + [C(qI - A_K)^{-1}B_K + D]u_k + e_k \end{aligned} \quad (2.48)$$

using $e_k = Fe_k^*$ where e_k^* has an identity covariance matrix, and defining

$$\tilde{y}_k = F^{-1}[I - C(qI - A_K)^{-1}K]y_k \quad (2.49a)$$

$$G(q) = F^{-1}C(qI - A_K)^{-1} \quad (2.49b)$$

$$D^* = F^{-1}D \quad (2.49c)$$

we obtain,

$$\begin{aligned} \tilde{y}_k &= G(q)B_K u_k + D^* u_k + G(q)x_0\delta_k + e_k^* \\ &= G(q) \otimes u_k^T \text{vec}(B_K) + I_{n_y} \otimes u_k^T \text{vec}(D^*) \\ &\quad + G(q)x_0\delta_k + e_k^* \end{aligned} \quad (2.50)$$

where $\text{vec}(B_K)$ and $\text{vec}(D^*)$ are vectorized B_K and D^* matrices along the rows. δ_k is the Kronecker delta function. Now $\text{vec}(B_K)$, $\text{vec}(D^*)$ and x_0 can be estimated using least squares from the above equation. The B , D matrices can be backed out as:

$$\hat{D} = F\hat{D}^* \quad (2.51a)$$

$$\hat{B} = \hat{B}_K + K\hat{D} \quad (2.51b)$$

2.4 Simulation and industrial case studies

In this section, the results of two simulation cases and an industrial case are reported to demonstrate the efficiency of proposed PARSIMs with comparison to N4SID in the System Identification Toolbox (Version 5.0) of Matlab. The first simulation is a second order single input and single output (SISO) counter example from [31]. The second is a Monte-Carlo simulation study over randomly chosen fourth order systems with two inputs and two outputs. The industrial case study is a 3×3 four-stage evaporator from [65].

2.4.1 Simulation example 1

The counter-example proposed in [31] is used here to test the effectiveness of the proposed parallel PARSIM methods.

$$x_{k+1} = \begin{bmatrix} 2\gamma & -\gamma^2 \\ 1 & 0 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ -2 \end{bmatrix} u_k + \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} e_k \quad (2.52a)$$

$$y_k = \begin{bmatrix} 2 & -1 \end{bmatrix} x_k + e_k \quad (2.52b)$$

where the variance of the noise process $var(e_k) = 217.1$, $\gamma = 0.9184$, $k_1 = -0.21$ and $k_2 = -0.559$ are used here. The system input is a high pass filter with unit white Gaussian noise as input.

$$u_k = (1 - \gamma q^{-1})^2 (1 + \gamma q^{-1})^2 \epsilon_k$$

For comparison we use the N4SID routine in Matlab, which actually implemented the CVA weighting, as the standard SIM algorithm. PEM implemented as the ARMAX routine in Matlab's System Identification Toolbox

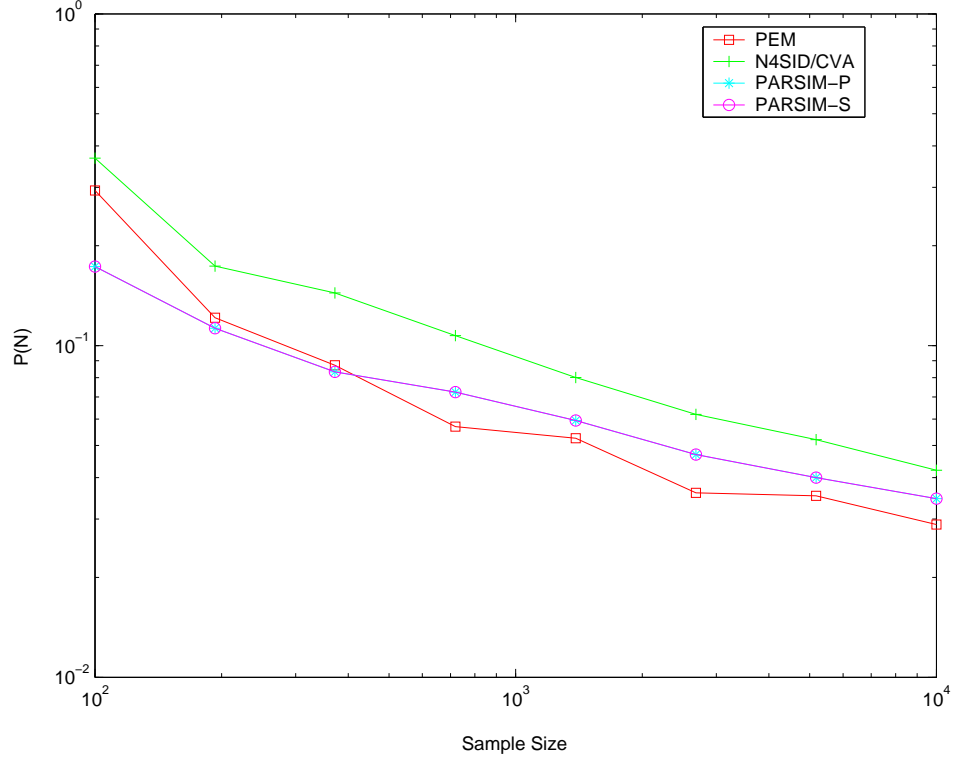


Figure 2.1: Asymptotic pole estimation results of the SISO counter example

is used as a benchmark. The performance of the methods is investigated with two indices, the standard deviation of the pole estimation errors and that of the zero estimation errors,

$$P(N) = \frac{1}{M} \sum_{k=1}^M \|\hat{P}_N^k - P_0\|_2 \quad (2.53a)$$

$$Z(N) = \frac{1}{M} \sum_{k=1}^M \|\hat{Z}_N^k - Z_0\|_2 \quad (2.53b)$$

where $M = 200$ is the number of independent simulations. \hat{P}_N^k and \hat{Z}_N^k are the estimated poles and zeros with N samples at k^{th} simulation, respectively.

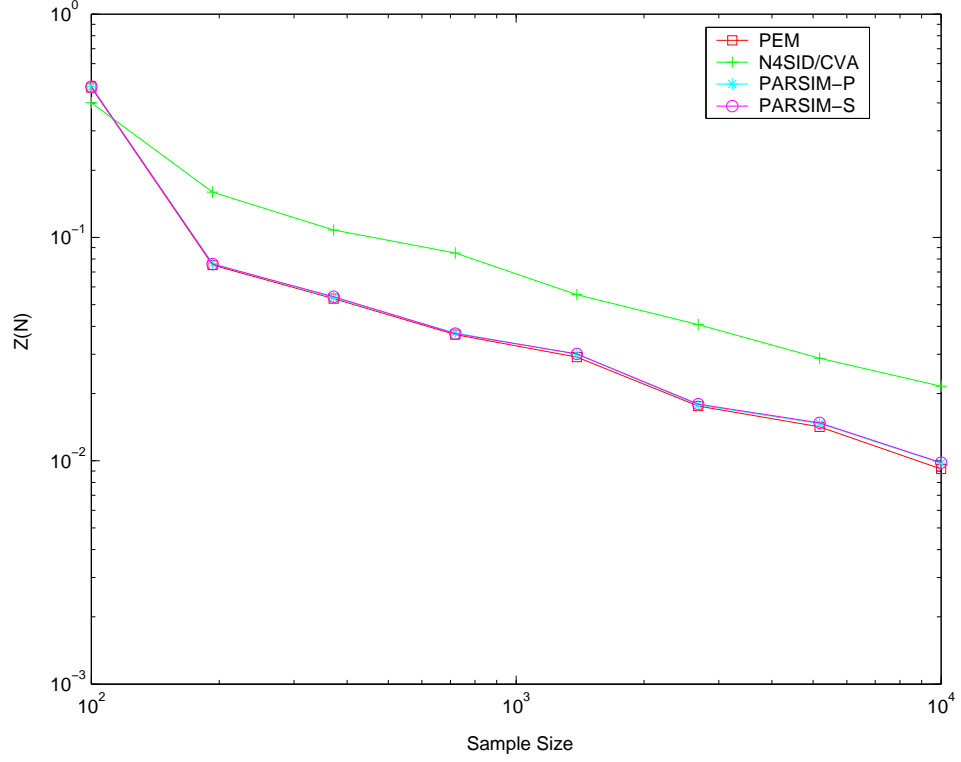


Figure 2.2: Asymptotic zero estimation results of the SISO counter example

P_0 and Z_0 are the true poles and zeros of the system, respectively. We choose $p = 7$, $f = 5$ for PARSIMs. The results of the simulations are shown in Fig. 2.1 and Fig. 2.2, which show the asymptotical performance of different algorithms. The results show that the PARSIMs outperform N4SID for both pole and zero estimation, and the zero estimates of PARSIMs are very close to those of PEM.

2.4.2 Simulation example 2

To study the potential benefits of the causal parameterization in PARSIMs, we perform a Monte-Carlo study over randomly chosen fourth order systems with two inputs and two outputs, estimated with the different methods. Since the motivation for the causal parameterization is to provide a better estimate of the observability matrix, we concentrate on the estimates of the A -matrix, viz. its eigenvalues. The input was chosen as a random binary signal with power up to 0.1 of the Nyquist frequency, and normal white noise with 0.1 times the unit covariance matrix was added to the output.

The system and the input/output data were generated by Matlab as follows

```
m0 = idss(drss(4,2,2));  
m0.d = zeros(2,2);  
m0.b = 5*randn(4,2);  
u = idinput([400,2], 'rbs', [0 0.1]);  
y = sim(m0,u) + 0.1*randn(400,2);
```

For each data set a model was estimated using Matlab's standard N4SID/CVA routine as well as using PARSIM-S and PARSIM-P. The future horizon (f) was chosen as 20 and the past horizon (p) was chosen as 10 in all cases.

For each method the standard deviations (absolute values) of each of

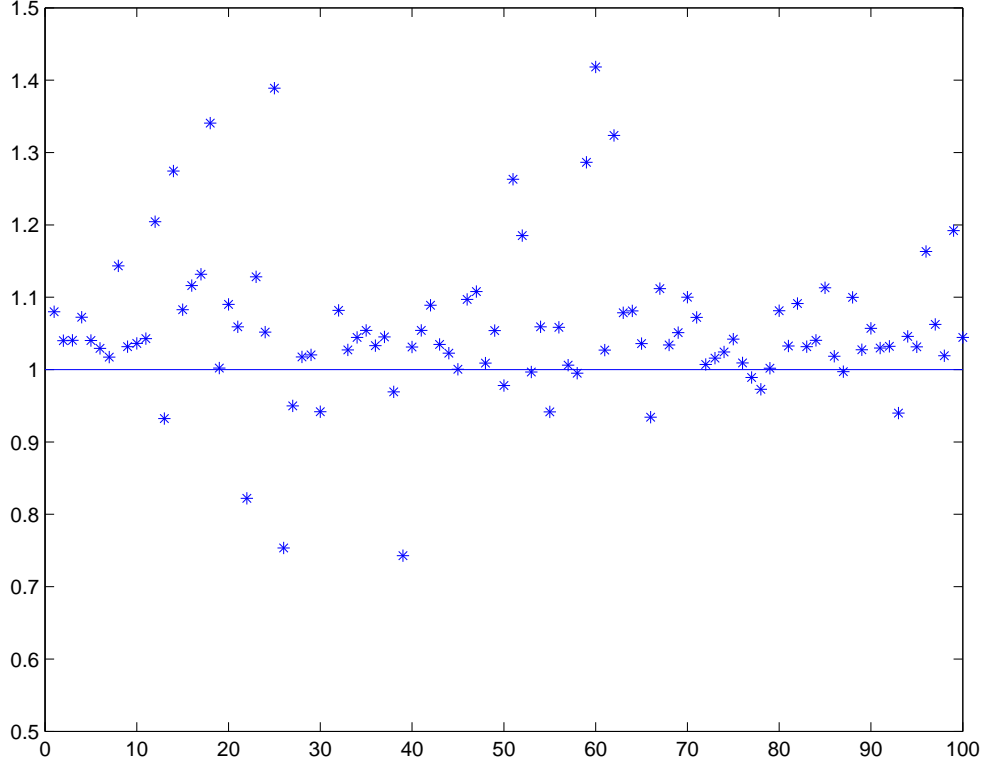


Figure 2.3: The value r defined by (2.54) for the 100 randomly chosen systems as defined in the text. The number of realizations for each system, M , was 25. (The value for system 31 is 4.46 and out of range.) PARSIM-P is better than N4SID/CVA in 84 of the 100 cases. The average excess of standard deviation for N4SID/CVA is 9.1%.

the four eigenvalues were estimated over the M realizations in the usual way:

$$\hat{\sigma}_i^{\text{CVA}}, \quad \hat{\sigma}_i^{\text{PARSIM-P}}, \quad i = 1, 2, 3, 4$$

As PARSIM-S and PARSIM-P gave nearly the same results, we only compare the performance of N4SID/CVA and PARSIM-P. The mean of the ratios of

the accuracy of the methods was computed

$$r = \frac{1}{4} \sum_{i=1}^4 \frac{\hat{\sigma}_i^{\text{CVA}}}{\hat{\sigma}_i^{\text{PARSIM-P}}} \quad (2.54)$$

as a measure of the relative accuracy of the two methods in estimating the poles/eigenvalues. A plot of r over the 100 different systems with $M = 25$ is given in Figure 2.3. It is seen that the enforcement of the causal model when estimating the observability matrix gives a noticeable improvement in the standard deviation of the eigenvalue estimates.

2.4.3 Industrial case study

In this subsection, the experimental data from a four-stage evaporator are analyzed [65]. The three inputs are feed flow, vapor flow to the first evaporator stage and cooling water flow. The three outputs are the dry matter content, the flow and temperature of the outcoming product. The time series plot of the data indicates that the inputs are PRBS. There are 6305 experimental data points, we use the first 3152 points for estimation and the rest of them for validation. We choose $p = 30$, $f = 20$ for PARSIMs and N4SID. By using the Akaike information criterion (AIC) and examining the singular values an 11th order system is chosen.

The coefficient of determination

$$R^2 = \left(1 - \frac{\sum_i [y_k(i) - \hat{y}_k(i)]^2}{\sum_i [y_k(i) - \bar{y}_k]^2}\right) \times 100 \quad (2.55)$$

of validation data is used as the metric for comparing different SIMs, where y_k , \hat{y}_k and \bar{y}_k are the measured output, simulated or predicted model output and

the mean of the output for the k^{th} output variables, respectively. The result of simulation and various horizons of prediction for different SIMs is shown in Table 2.1.

From the result, we can see that all methods work well for one-step ahead predictions. As the prediction horizon increases, the prediction accuracy decreases, as expected. In general the PARSIM algorithms outperform the N4SID on long term predictions. For simulation error N4SID failed on y_3 . While the N4SID results are almost the same as the PARSIM results for 1 and 20 steps ahead predictions, the PARSIMs produce better results for 100 steps ahead prediction and simulation.

2.5 Summary

In this chapter, a novel subspace identification approach is proposed to enforce the causality of high order ARX models. The key idea is to avoid the estimation of parameters that are known to be zero. This means that a lower triangular structure of an estimated matrix must be enforced which leads to somewhat more complicated calculations. Also other authors have noted the potential problems that arise from these non-causal elements. Ljung and McKelvey [46] have noted that the problems with closed loop data have their roots in these non-causal terms. They suggest to use explicitly computed k -step ahead predictions from a single causal ARX-model. This is very different from the algorithm suggested in this paper, and apparently it does not make the best use of the observed data. The new algorithms, which fall into the

subspace fitting framework, are shown to be consistent under mild assumptions and applicable to a general state space model structure.

	1 step ahead	20 steps ahead	100 steps ahead	Simulation	
y_1	74.79	60.16	49.64	44.35	N4SID
	74.65	61.20	54.14	51.24	PARSIM-P
	74.52	60.75	53.87	48.16	PARSIM-S
y_2	62.12	60.27	60.12	58.15	N4SID
	61.62	60.07	60.28	59.97	PARSIM-P
	61.43	60.43	60.01	59.55	PARSIM-S
y_3	84.50	57.27	14.27	—	N4SID
	84.47	60.60	42.06	30.35	PARSIM-P
	84.47	60.64	34.49	15.35	PARSIM-S

Table 2.1: The model fit as measured by R^2 in (2.55) of identified models for simulation and prediction of validation data for the evaporator. (R^2 less than zero is indicated by '—'.)

We have shown that the variance of the observability matrix estimates is in general smaller if the non-causal terms are omitted. It is difficult to make further comparison about the variance of the system matrices because they depend on the basis. Simulation tests are conducted to compare the variance of the eigenvalues of the A matrix. We have indeed seen improved behavior in the reported tests. The simulation studies indicate that the proposed algorithms are superior to SIMs with CVA weighting, which are considered optimal.

Chapter 3

Closed-loop Subspace Identification with Innovation Estimation

3.1 Introduction

The closed-loop identification is of special interest for a large number of engineering applications. For safety reasons and quality restrictions, it is desirable that identification experiments are carried out under the closed-loop or partial closed-loop condition. As pointed out by many researchers [45, 62], the fundamental problem with closed-loop data is the correlation between the unmeasurable noise and the input. Traditional closed-loop identification approaches fall into the prediction error methods (PEMs) framework. A comprehensive study in this area is provided by Forssell and Ljung [17]. Based on their analysis, the closed-loop identification methods can be categorized into three main groups: the direct approach, the indirect approach, and the joint input-output approach.

Contrary to the open loop SIMs, the traditional SIMs (e.g., CVA, N4SID and MOESP) are biased under closed-loop condition. Verhaegen [66] proposed a closed-loop SIM via the identification of an overall open-loop state space model followed by a model reduction step to obtain state space represen-

tations of the plant and controller. The disadvantages of the approach is that a high order overall system has to be identified, which introduces extra computational burden. Ljung and McKelvey [46] investigated the SIM through the classical realization path and proposed a recursive approach based on ARX model as a feasible closed-loop SIM. The drawback of the approach is that the ARX parametrization is not applicable for the generic system. Recently, Chiuso and Picci [10] analyzed SIMs with feedback through stochastic realization theory and provided a theoretical analysis to construct the geometric state based on an oblique predictor space. Nevertheless, they did not provide any algorithm in detail.

Formulated in an errors-in-variables (EIV) framework, Chou and Verhaegen [11] proposed a new SIM that can be applied to closed-loop data. The algorithm is nevertheless very complex which has to treat the case of white input from non-white input differently. Wang and Qin [69] proposed the use of parity space and principal component analysis (PCA) for EIV and closed-loop identification which is applicable to correlated input excitation. Huang, Ding and Qin [27] analyzed the reason why these methods cannot be applied to white input directly and proposed a revised instrumental variables approach.

To the best of our knowledge, the possibility of closed-loop identification with SIMs has not been thoroughly analyzed. The main purpose of this chapter is to reveal the feasibility of the consistent estimation with SIMs under the closed-loop operation. It is shown that the consistency of closed-loop SIMs can be achieved through innovation estimation.

The rest of the chapter is organized as follows. In Section 3.2, we analyze feasibility of closed-loop SIMs through innovation estimation. The consistency of closed-loop SIMs is also presented in this section. Based on this analysis, two feasible closed-loop SIMs are presented in detail in Section 3.3. In Section 3.4, we compare the "innovation estimation" approach with the "whitening filter" approach. The similarity and difference between them are investigated in detail. Section 3.5 concludes the chapter.

3.2 Analysis of subspace identification under closed-loop condition

3.2.1 Problem formulation and assumptions

In this work, we assume that the system to be identified can be written in innovations form as

$$x_{k+1} = Ax_k + Bu_k + Ke_k \quad (3.1a)$$

$$y_k = Cx_k + Du_k + e_k \quad (3.1b)$$

where $y_k \in \Re^{n_y}$, $x_k \in \Re^n$, $u_k \in \Re^{n_u}$, and $e_k \in \Re^{n_y}$ are the system output, state, input, and innovation, respectively. A , B , C and D are system matrices with appropriate dimensions. K is the Kalman filter gain. The system described by (3.1) can also be presented as

$$y_k = G(q)u_k + H(q)e_k \quad (3.2a)$$

where $G(q) = C(qI - A)^{-1}B + D$, and $H(q) = C(qI - A)^{-1}K + I$. We shall assume that the input is determined through feedback as

$$u_k = -F(q)y_k + r_k \quad (3.3)$$

where r_k is the reference signal, and $F(q)$ is the filter standing for the feedback mechanism.

To establish the statistical consistency of the SIM under closed-loop condition, we introduce following assumptions:

A1 : The eigenvalues of $A - KC$ are strictly inside the unit circle.

A2 : The system is minimal in the sense that (C, A) is observable and $(A, [B, K])$ is controllable.

A3 : The innovation sequences e_k is a stationary, zero mean, white noise process with the second order moments

$$E(e_i e_j^T) = R \delta_{ij}$$

where δ_{ij} is the Kronecker delta.

A4 : The input $u(k)$ and innovation sequence $e(j)$ are uncorrelated for $\forall j \geq k$, which implies that either the system or the controller contains a delay.

A5 : The reference signal and innovation sequence are uncorrelated to each other, and the reference signal is persistently exciting of a sufficient order.

A6 : The closed-loop subsystem from r and e to y are asymptotically stable.

A7 : Either the past horizon $p \rightarrow \infty$ or $A - KC$ is nilpotent.

The closed-loop identification problem is: given a set of input/output measurements and reference measurements, estimate the system matrices, Kalman filter gain K up to within a similarity transformation, and the innovation covariance matrix R . The exact knowledge of the controller is not required for the closed-loop identification approach proposed in this work.

Based on state space description in (3.1), an extended state space model can be formulated as

$$Y_f = \Gamma_f X_k + H_f U_f + G_f E_f \quad (3.4a)$$

$$Y_p = \Gamma_p X_{k-p} + H_p U_p + G_p E_p \quad (3.4b)$$

where the subscripts f and p denote future and past horizons, respectively.

The extended observability matrix is

$$\Gamma_f = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{f-1} \end{bmatrix} \quad (3.5)$$

and H_f and G_f are Toeplitz matrices:

$$H_f = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-2}B & CA^{f-3}B & \cdots & D \end{bmatrix} \quad (3.6a)$$

$$G_f = \begin{bmatrix} I & 0 & \cdots & 0 \\ CK & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-2}K & CA^{f-3}K & \cdots & I \end{bmatrix} \quad (3.6b)$$

The input and output data are arranged in the following Hankel form:

$$U_f = \begin{bmatrix} u_k & u_{k+1} & \cdots & u_{k+N-1} \\ u_{k+1} & u_{k+2} & \cdots & u_{k+N} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k+f-1} & u_{k+f} & \cdots & u_{k+f+N-2} \end{bmatrix} \quad (3.7a)$$

$$U_p = \begin{bmatrix} u_{k-p} & u_{k-p+1} & \cdots & u_{k-p+N-1} \\ u_{k-p+1} & u_{k-p+2} & \cdots & u_{k-p+N} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k-1} & u_k & \cdots & u_{k+N-2} \end{bmatrix} \quad (3.7b)$$

The state sequences are defined as:

$$X_k = [x_k, x_{k+1}, \cdots, x_{k+N-1}] \quad (3.8a)$$

$$X_{k-p} = [x_{k-p}, x_{k-p+1}, \cdots, x_{k-p+N-1}] \quad (3.8b)$$

Similar formulations are made for Y_f , Y_p , E_f , and E_p . Subspace identification consists of estimating the extended observability matrix first and then the model parameters.

3.2.2 Analysis of the closed-loop SIM

The main purpose of the subsection is to explore the feasibility of closed-loop SIMs with innovation estimation. We can partition the extended state space model in (3.4) row-wise as follows:

$$Y_f = \begin{bmatrix} Y_{f1} \\ Y_{f2} \\ \vdots \\ Y_{ff} \end{bmatrix}; \quad Y_i \triangleq \begin{bmatrix} Y_{f1} \\ Y_{f2} \\ \vdots \\ Y_{fi} \end{bmatrix}; \quad i = 1, 2, \dots, f \quad (3.9)$$

Partition U_f and E_f in a similar way to define U_{fi} , U_i , E_{fi} , and E_i , respectively, for $i = 1, 2, \dots, f$. Denote further

$$\Gamma_f = \begin{bmatrix} \Gamma_{f1} \\ \Gamma_{f2} \\ \vdots \\ \Gamma_{ff} \end{bmatrix} \quad (3.10a)$$

$$H_{fi} \triangleq \begin{bmatrix} CA^{i-2}B & \cdots & CB & D \end{bmatrix} \quad (3.10b)$$

$$\triangleq \begin{bmatrix} H_{i-1} & \cdots & H_1 & H_0 \end{bmatrix} \quad (3.10c)$$

$$G_{fi} \triangleq \begin{bmatrix} CA^{i-2}K & \cdots & CK & I \end{bmatrix} \quad (3.10d)$$

$$\triangleq \begin{bmatrix} G_{i-1} & \cdots & G_1 & G_0 \end{bmatrix} \quad (3.10e)$$

where H_i and G_i are the Markov parameters for the deterministic input and innovation sequence, respectively. We have the following partitioned equations:

$$Y_{fi} = \Gamma_{fi}X_k + H_{fi}U_i + G_{fi}E_i \quad (3.11)$$

for $i = 1, 2, \dots, f$. Denote further,

$$H_{fi}^- \triangleq \begin{bmatrix} H_{i-1} & \cdots & H_1 \end{bmatrix} \quad (3.12a)$$

$$G_{fi}^- \triangleq \begin{bmatrix} G_{i-1} & \cdots & G_1 \end{bmatrix} \quad (3.12b)$$

The partitioned Y_{fi} in (3.11) is equal to

$$Y_{fi} = \Gamma_{fi}X_k + H_{fi}^-U_{i-1} + H_{f1}U_1 + G_{fi}^-E_{i-1} + E_{fi} \quad (3.13)$$

By eliminating e_k in the innovation model (3.1) through iteration, it is straightforward to derive the following relation,

$$X_k = L_z Z_p + A_K^p X_{k-p} \quad (3.14)$$

where

$$L_z \triangleq \begin{bmatrix} \Delta_p(A_K, K) & \Delta_p(A_K, B_K) \end{bmatrix} \quad (3.15a)$$

$$\Delta_p(A, B) \triangleq \begin{bmatrix} A^{p-1}B & \cdots & AB & B \end{bmatrix} \quad (3.15b)$$

$$A_K \triangleq A - KC \quad (3.15c)$$

$$B_K \triangleq B - KD \quad (3.15d)$$

$$Z_p \triangleq \begin{bmatrix} Y_p^T & U_p^T \end{bmatrix}^T \quad (3.15e)$$

Substituting this equation into (3.13), we obtain

$$Y_{fi} = \Gamma_{fi} L_z Z_p + \Gamma_{fi} A_K^p X_{k-p} + H_{fi}^- U_{i-1} + H_{f1} U_1 + G_{fi}^- E_{i-1} + E_{fi} \quad (3.16)$$

for $i = 1, 2, \dots, f$. Note that the second term in the right hand side (RHS) of (3.16) tends to zero as p tends to infinity under assumption A1. To facilitate the derivation of the main results, we assume that, in this subsection, the process described by (3.1) does not contain the direct term, i.e., $H_{f1} = 0$. Therefore, (3.16) reduced to

$$Y_{fi} = \Gamma_{fi} L_z Z_p + H_{fi}^- U_{i-1} + G_{fi}^- E_{i-1} + E_{fi} \quad (3.17)$$

Since the future innovation, E_{fi} , is uncorrelated with Z_p , U_{i-1} and E_{i-1} in (3.17) under closed-loop condition. If the E_f is already known, the coefficient

matrices can be estimated through a straightforward linear regression as:

$$\begin{bmatrix} \hat{\Gamma}_{fi}L_z & \hat{H}_{fi}^- & \hat{G}_{fi}^- \end{bmatrix} = Y_{fi} \begin{bmatrix} Z_p \\ U_{i-1} \\ E_{i-1} \end{bmatrix}^\dagger \quad (3.18)$$

where \dagger stands for pseudo-inverse operation. A remaining question is whether this approach gives us a consistent estimate of $\Gamma_{fi}L_z$, H_{fi}^- and G_{fi}^- . Here we formulate the results as follows.

Lemma 3.2.1. *Under the assumptions stated in Subsection 2.1, the joint input-output signal, $\chi_k = \begin{bmatrix} y_k \\ u_k \end{bmatrix}$, is persistently exciting of any order.*

[**Proof**] See Appendix A10.1 in [62].

Theorem 3.2.2. *The estimates of $\Gamma_{fi}L_z$, H_{fi}^- and G_{fi}^- in (3.18) are consistent for $\forall i = 1, 2, \dots, f$ if and only if the joint input-output signal, $\chi_k = \begin{bmatrix} y_k \\ u_k \end{bmatrix}$, is persistently exciting to the order of $p + f - 1$, where p and f denote the past and the future horizons, respectively.*

[**Proof**] See Appendix A.

[**Remark 1**] The analysis of consistency for the case of $D \neq 0$ is similar to that presented in this subsection with the help of the assumption A4, while it requires χ_k to be consistently exciting to the order of $f + p$. The proof is similar to the one provided in Appendix A through a minor modification.

[**Remark 2**] From the derivation in Appendix A, we can see that the key is to maintain the full row rank of $\begin{bmatrix} Z_p^T & U_{i-1}^T & E_{i-1}^T \end{bmatrix}^T$, which is the assumption for the existence of the oblique predict space in Chiuso and Picci's [10] analysis.

3.3 Closed-loop subspace identification methods with innovation estimation

From the analysis in Section 3.2, we can conclude that, under certain assumptions, the consistency estimation with SIMs can be achieved if the innovation sequence is already known. The only challenge left now is how to estimate the innovation signal. Qin and Ljung [58] provided an algorithm for closed-loop SIM with innovation estimation method (PARSIM-E). In this section, we introduce the algorithm in more detail and provide another variation of it (PARSIM-E1).

3.3.1 Algorithm 1: PARSIM-E

In this subsection, we present a closed-loop SIM algorithm with innovation estimation. Similar to Subsection 3.2.2, in this subsection, we assume that $D = 0$ in (3.1). By ignoring the second term on the RHS of (3.16) and set $i = 1$, we have

$$Y_{f1} = \Gamma_{f1} L_z Z_p + E_1 \quad (3.19)$$

Therefore, a least squares estimate of the innovation process is:

$$\hat{E}_1 = Y_{f1} - \hat{\Gamma}_{f1} L_z Z_p \quad (3.20)$$

where

$$\hat{\Gamma}_{f1} L_z = Y_{f1} Z_p^\dagger \quad (3.21)$$

Now return to (3.17) for a general $i = 2, 3, \dots, f$. Noting that

$$E_i = \begin{bmatrix} E_{f1} \\ E_{f2} \\ \vdots \\ E_{fi} \end{bmatrix} = \begin{bmatrix} E_{i-1} \\ E_{fi} \end{bmatrix} \quad (3.22)$$

and replacing E_{i-1} with \hat{E}_{i-1} , (3.17) becomes

$$Y_{fi} = \begin{bmatrix} \Gamma_{fi} L_z & H_{fi}^- & G_{fi}^- \end{bmatrix} \begin{bmatrix} Z_p \\ U_{i-1} \\ \hat{E}_{i-1} \end{bmatrix} + E_{fi} \quad (3.23)$$

The least squares estimate

$$\begin{bmatrix} \hat{\Gamma}_{fi} L_z & \hat{H}_{fi}^- & \hat{G}_{fi}^- \end{bmatrix} = Y_{fi} \begin{bmatrix} Z_p \\ U_{i-1} \\ \hat{E}_{i-1} \end{bmatrix}^\dagger \quad (3.24)$$

With the least squares estimates of $\Gamma_{fi} L_z$ from (3.21) and (3.34), we obtain

$$\hat{\Gamma}_f L_z = \begin{bmatrix} \hat{\Gamma}_{f1} L_z \\ \hat{\Gamma}_{f2} L_z \\ \vdots \\ \hat{\Gamma}_{ff} L_z \end{bmatrix}$$

The observability matrix, Γ_f , can be estimated similarly to the PARSIM-E procedures given in [59].

$$\hat{\Gamma}_f = W_1^{-1} U S^{1/2} \quad (3.25)$$

where

$$W_1(\hat{\Gamma}_f L_z) W_2 = U S V^T$$

and the weighting matrices

$$W_1 = I \quad (3.26a)$$

$$W_2 = (Z_p \Pi_{U_f}^\perp Z_p^T)^{1/2} \quad (3.26b)$$

Theorem 3.3.1. *The estimate of $\hat{\Gamma}_f$ from PARSIM-E is consistent under the assumptions A1 to A7 stated in Section 3.2.*

[Proof] To prove the consistency of $\hat{\Gamma}_f$ from PARSIM-E, it is sufficient to show that as $N \rightarrow \infty$

$$\begin{bmatrix} \hat{\Gamma}_{fi} L_z & \hat{H}_{fi}^- & \hat{G}_{fi}^- \end{bmatrix} \rightarrow \begin{bmatrix} \Gamma_{fi} L_z & H_{fi}^- & G_{fi}^- \end{bmatrix}$$

Note that if the innovation sequence is already known it has been proven in Theorem 3.2.2. Therefore, Theorem 3.3.2 is valid if

$$\hat{E}_{i-1} \rightarrow E_{i-1}$$

as $N \rightarrow \infty$, which can be proven recursively with the help of Theorem 3.2.2.

[Remark 1] After obtaining the consistent estimate of the extended observability matrix, the A and C matrices can be estimated as in [67].

[Remark 2] For $D \neq 0$ case, the PARSIM-E is consistent as well if there is a delay in the controller.

3.3.2 Algorithm 2: PARSIM-E1

In this subsection, we present another closed-loop SIM algorithm with innovation estimation.

By ignoring the second term on the RHS of (3.16) and set $i = 1$, we have

$$Y_{f1} = \Gamma_{f1} L_z Z_p + E_1 \quad (3.27)$$

Furthermore, if we set the future horizon, $f = 1$,

$$Y_{11} = \Gamma_{11} L_z Z_p + E_{11} \quad (3.28)$$

where Y_{11} is defined in (3.9) and E_{11} is defined in a similar way. Γ_{11} is defined in (3.10a).

Therefore, a least squares estimate of the innovation process is:

$$\hat{E}_{11} = Y_{11} - \hat{\Gamma}_{11} L_z Z_p \quad (3.29)$$

where

$$\hat{\Gamma}_{11} L_z = Y_{11} Z_p^\dagger \quad (3.30)$$

After obtaining estimates of the innovation sequence, it is straightforward to construct \hat{E}_f based on

$$\hat{E}_f = \begin{bmatrix} \hat{e}_k & \hat{e}_{k+1} & \cdots & \hat{e}_{k+N-1} \\ \hat{e}_{k+1} & \hat{e}_{k+2} & \cdots & \hat{e}_{k+N} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{e}_{k+f-1} & \hat{e}_{k+f} & \cdots & \hat{e}_{k+f+N-2} \end{bmatrix} \quad (3.31)$$

Now return to (3.17) for a general $i = 1, 2, \dots, f$. Noting that

$$E_i = \begin{bmatrix} E_{f1} \\ E_{f2} \\ \vdots \\ E_{fi} \end{bmatrix} = \begin{bmatrix} E_{i-1} \\ E_{fi} \end{bmatrix} \quad (3.32)$$

and replacing E_{i-1} with \hat{E}_{i-1} , (3.17) becomes

$$Y_{fi} = \begin{bmatrix} \Gamma_{fi}L_z & H_{fi}^- & G_{fi}^- \end{bmatrix} \begin{bmatrix} Z_p \\ U_{i-1} \\ \hat{E}_{i-1} \end{bmatrix} + E_{fi} \quad (3.33)$$

The least squares estimate

$$\begin{bmatrix} \hat{\Gamma}_{fi}L_z & \hat{H}_{fi}^- & \hat{G}_{fi}^- \end{bmatrix} = Y_{fi} \begin{bmatrix} Z_p \\ U_{i-1} \\ \hat{E}_{i-1} \end{bmatrix}^\dagger \quad (3.34)$$

With the least squares estimates of $\Gamma_{fi}L_z$ from (3.34), we obtain

$$\hat{\Gamma}_f L_z = \begin{bmatrix} \hat{\Gamma}_{f1}L_z \\ \hat{\Gamma}_{f2}L_z \\ \vdots \\ \hat{\Gamma}_{ff}L_z \end{bmatrix} \quad (3.35)$$

The observability matrix, Γ_f , can be estimated similarly to PARSIM-E.

Theorem 3.3.2. *The estimate of $\hat{\Gamma}_f$ from PARSIM-E1 is consistent under the assumptions A1 to A7 stated in Section 3.2.*

[Proof] To prove the consistency of $\hat{\Gamma}_f$ from PARSIM-E1, it is sufficient to show that as $N \rightarrow \infty$

$$\begin{bmatrix} \hat{\Gamma}_{fi}L_z & \hat{H}_{fi}^- & \hat{G}_{fi}^- \end{bmatrix} \rightarrow \begin{bmatrix} \Gamma_{fi}L_z & H_{fi}^- & G_{fi}^- \end{bmatrix}$$

Note that if the innovation sequence is already known it has been proven in Theorem 3.2.2. Therefore, Theorem 3.3.2 is valid if

$$\hat{E}_{11} \rightarrow E_{11}$$

as $N \rightarrow \infty$, which is straightforward for a sufficient large past horizon.

3.3.3 K estimation under closed-loop condition

De Ruscio [14] introduced a method to identify the Kalman filter gain with QR implementation for open loop data. It requires that E_f is uncorrelated to Z_p and U_f , which is invalid under closed-loop condition. In this subsection, we provide a new way to calculate K with closed-loop data.

After the determination of the system order through the Akaike information criterion, we can obtain the estimate of the innovation sequence e_k , which can be used to construct \hat{E}_f . Substituting the X_k in the extended state space model (3.4) with (3.14), we obtain

$$Y_f = \Gamma_f L_z Z_p + \Gamma_f A_K^p X_{k-p} + H_f U_f + G_f E_f \quad (3.36)$$

Omitting the second term for a sufficient large p , and replacing $\Gamma_f L_z$ with $\hat{\Gamma}_f L_z$, (3.36) becomes

$$\tilde{Y}_f = H_f U_f + G_f E_f \quad (3.37)$$

where

$$\tilde{Y}_f = Y_f - \hat{\Gamma}_f L_z Z_p$$

We can partition \tilde{Y}_f row-wise as follows:

$$\tilde{Y}_{fi} = H_{fi} U_i + G_{fi}^- E_{i-1} + E_{fi}; \quad i = 1, 2, \dots, f \quad (3.38)$$

Again, replacing E_{i-1} with \hat{E}_{i-1} , the least squares estimation

$$\begin{bmatrix} \hat{H}_{fi} & \hat{G}_{fi}^- \end{bmatrix} = \tilde{Y}_{fi} \begin{bmatrix} U_i \\ \hat{E}_{i-1} \end{bmatrix}^\dagger; \quad i = 2, 3, \dots, f \quad (3.39)$$

With the definition of G_{fi}^- in (3.52c),

$$\hat{G}_{fi} = \begin{bmatrix} \hat{G}_{fi}^- & I \end{bmatrix}; i = 2, 3, \dots, f \quad (3.40)$$

where \hat{G}_{fi} is the estimation of G_{fi} in (3.10d).

Therefore, the estimate of G_f can be obtained based on (3.6b), which is lower triangular but is not exactly Toeplitz due to estimation error. After taking the average of the diagonal block components of \hat{G}_f , the Toeplitz structure of \hat{G}_f can be preserve as,

$$\hat{G}_f = \begin{bmatrix} I & 0 & \cdots & 0 \\ \hat{G}_1 & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{G}_{f-1} & \hat{G}_{f-2} & \cdots & I \end{bmatrix} \quad (3.41)$$

Furthermore, notice that

$$\begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_{f-1} \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{f-2} \end{bmatrix} K = \Gamma_{f-1} K$$

The Kalman gain, K , can be calculated as

$$\hat{K} = \hat{\Gamma}_{f-1}^\dagger \begin{bmatrix} \hat{G}_1 \\ \hat{G}_2 \\ \vdots \\ \hat{G}_{f-1} \end{bmatrix} \quad (3.42)$$

where $\hat{\Gamma}_{f-1}^\dagger$ can be obtained as discussed in Subsection 3.3.1. To make the eigenvalue of the predictor $A - KC$ lie strictly inside the unit circle, the \hat{K} can be further refined by solving the steady state algebraic Riccati equation [2].

After calculating the estimates of K , the B and D matrices can be estimated optimally using the estimates of A , C , K and F [59].

3.3.4 Simulation studies

In this subsection, two simulated case studies are reported under closed-loop condition. For comparison we use the N4SID routine in the System Identification Toolbox (Version 5.0) of Matlab, which actually implemented the canonical variate analysis (CVA) weighting, as the standard SIM algorithm.

3.3.4.1 Simulation example: a SISO process

We simulate the following single input and single output (SISO) process

$$y_k + ay_{k-1} = bu_{k-1} + e_k + ce_{k-1} \quad (3.43)$$

where $a = -0.9$, $b = 1$, and $c = 0.5$. The feedback controller is

$$u_k = -Ky_k + r_k \quad (3.44)$$

The reference signal, r_k , and innovation sequence, e_k , are white noise with $cov(r_k) = 2$, and $cov(e_k) = 1$, respectively. Open-loop experiments are simulated with $K = 0$ and closed-loop experiments with $K = 0.6$. In both cases, we choose $p = 9$, $f = 3$ for PARSIM-Es, and run 20 independent Monte Carlo simulations. The sampling size for each experiment is 4000. The identification results from PEM implemented using the ARMAX routine in Matlab's System Identification Toolbox are used here as a benchmark.

The pole estimation results for both open-loop and closed-loop experiments are shown in Fig. 3.1. From the results, we can conclude that the performances of all three methods are excellent in the open-loop case. For the close-loop identification, the estimate from PARSIM-E is comparable with that from PEM, while the traditional SIM with the CVA weighting fail to provide consistent estimates.

The results of frequency response estimations for the closed-loop simulation are shown in Fig. 3.2, Fig. 3.3, Fig. 3.4 and Fig. 3.5. From the results we can see that the estimate of frequency response from N4SID is biased. The identification results from PARSIM-E is very close to those from PEM.

3.3.4.2 Simulation example: a MIMO process

In this subsection, we simulate the following 2×2 linear dynamic system

$$x_{k+1} = \begin{bmatrix} 0.67 & 0.67 & 0 & 0 \\ -0.67 & 0.67 & 0 & 0 \\ 0 & 0 & -0.67 & -0.67 \\ 0 & 0 & 0.67 & -0.67 \end{bmatrix} x_k \quad (3.45a)$$

$$+ \begin{bmatrix} 0.6598 & -0.5256 \\ 1.9698 & 0.4845 \\ 4.3171 & -0.4879 \\ -2.6436 & -0.3416 \end{bmatrix} u_k$$

$$+ \begin{bmatrix} -0.6968 & -0.1474 \\ 0.1722 & 0.5646 \\ 0.6484 & -0.4660 \\ -0.9400 & 0.1032 \end{bmatrix} e_k$$

$$y_k = \begin{bmatrix} -0.3749 & 0.0751 & -0.5225 & 0.5830 \\ -0.8977 & 0.7543 & 0.1159 & 0.0982 \end{bmatrix} x_k + e_k \quad (3.45b)$$

The output feedback controller is

$$u_k = r_k + F_b y_k \quad (3.46)$$

where r_k is the reference signal and F_b is the feedback gain matrix. In the experiment, we use the pseudo-random binary signals (PRBS) with clock period of 5 samples as the reference sequences. 4000 samples of the input and output data are generated to identify the model with

$$\text{cov}(e_k) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$F_b = \begin{bmatrix} -0.25 & 0 \\ 0 & -0.25 \end{bmatrix}$$

We choose $p = 9$, $f = 5$ for PARSIM-Es, and run 10 independent Monte Carlo simulations.

The pole estimation results for the closed-loop experiments are shown in Fig. 3.6, Fig. 3.7 and Fig. 3.8. From the results we can see that the PARSIM-Es provide consistent estimates, while the N4SID subroutine with CVA weighting results in biased estimates.

The estimates of the frequency response for the closed-loop simulations are shown in Fig. 3.9, Fig. 3.10 and Fig. 3.11. We can see that the estimated frequency responses from PARSIM-Es match well with that of the real system. The traditional SIM fails to provide the consistent frequency responses.

3.4 Comparisons of closed-loop subspace identification methods

In this section, we introduce another closed-loop SIM denoted as "whitening filter" approach and investigate the similarity and difference of the approach and the "innovation estimation" approach.

Notice that the system described by (3.1) can also be represented as

$$x_{k+1} = A_K x_k + B_K u_k + K y_k \quad (3.47a)$$

$$y_k = C x_k + D u_k + e_k \quad (3.47b)$$

where $y_k \in \mathbb{R}^{n_y}$, $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^{n_u}$ and $e_k \in \mathbb{R}^{n_y}$ are the system output, state, input and innovation respectively. A and C are system matrices with appropriate dimensions. K is the Kalman filter gain. $A_K = A - KC$, and $B_K = B - KD$. We refer to (3.47) as the state space model in predictor form.

The system represented by (3.1) and the represented by (3.47) are equivalent, but system (3.1) uses the original process A matrix while system (3.47) uses the predictor A_K matrix. If the process to be identified is unstable, the predictor A_K matrix can still be stable. The closed-loop identification problem is: given a set of input/output and reference measurements under closed-loop, estimate the system matrices (A, B, C, D) and Kalman filter gain K up to a similarity transformation.

3.4.1 Closed-loop Subspace Identification Methods

3.4.1.1 Overview of closed-loop SIMs

Based on state space description in (3.1), an extended state space model can be formulated as (3.4)

$$Y_f = \Gamma_f X_k + H_f U_f + G_f E_f$$

Solving x_k by iterating (3.47), it is straightforward to derive the following relation as (3.14),

$$X_k = L_z Z_p + A_K^p X_{k-p}$$

Substituting (3.14) into (3.4), we obtain

$$Y_f = \Gamma_f A_K^p X_{k-p} + \Gamma_f L_z Z_p + H_f U_f + G_f E_f \quad (3.48)$$

If the past horizon p is large enough, the first term on the RHS tends to zero for stable A_K . The last two terms of the RHS of (3.48) are correlated for closed-loop systems. Therefore, most of the closed-loop SIMs try to decouple these two terms. The SIMPCA methods proposed in [69] and a later modification in [27] move $H_f U_f$ to the LHS and use principal component analysis on the joint input/output data simultaneously. We refer to these approaches as one-step approaches since no pre-estimation is needed. Another approach that falls in the one-step approach category is the observer/Kalman filter ID (OKID) by [52].

Since equation (3.48) is actually composed of f block rows in each term and the first block row gives an estimate of the innovation, Qin and Ljung [58]

propose an innovation estimation method (PARSIM-E) that partition (3.48) into f block rows and use the estimated innovation from previous block rows to further estimate model parameters of the next block row sequentially. An alternative method known as PARSIM-E1 [40] estimates the innovation from the first block row and then treats \hat{e}_k as known to estimate other model parameters. The SSARX approach proposed in [29] uses the predictor form (3.47) and pre-estimate a high order ARX model parameter to decouple the correlation between U_f and E_f . The well known CVA algorithm proposed by Larimore [37] actually pre-estimate H_f using a high order ARX and the move $\hat{H}_f U_f$ to the LHS of (3.48). Shi and MacGragor [61] also use this technique.

These approaches are referred to as two-step approaches in which a pre-estimation step is needed to decouple the noise and control input. The pre-estimation step is usually done by a high-order ARX; only different information is used to carry out the main step.

Inspired from the SSARX approach, Chiuso and Picci [9] give a variation known as the whitening filter approach that uses the predictor model form and carry out multi-stage projections row by row. In each block row projection causality is strictly enforced, similar to [57]. No pre-estimation is involved but the projections have to be done block-row wise to decouple noise from control input. We refer to these approaches as multi-stage projection approaches.

3.4.1.2 Whitening Filter Approach

Based on state space description in (3.47), an alternative extended state space model can be formulated as

$$Y_f = \bar{\Gamma}_f X_k + \bar{H}_f U_f + \bar{G}_f Y_f + E_f \quad (3.49)$$

The modified extended observability matrix is

$$\bar{\Gamma}_f = \begin{bmatrix} C \\ CA_K \\ \vdots \\ CA_K^{f-1} \end{bmatrix} \quad (3.50)$$

and \bar{H}_f and \bar{G}_f are:

$$\bar{H}_f = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB_K & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA_K^{f-2} B_K & CA_K^{f-3} B_K & \cdots & D \end{bmatrix} \quad (3.51a)$$

$$\bar{G}_f = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ CK & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA_K^{f-2} K & CA_K^{f-3} K & \cdots & 0 \end{bmatrix} \quad (3.51b)$$

Similar to the innovation estimation method, one can substitute (3.14) into (3.49) and partition the resulting equation row-wise as

$$Y_{fi} = \bar{\Gamma}_{fi} L_z Z_p + \bar{\Gamma}_{fi} A_K^p X_{k-p} + \bar{H}_{fi} U_i + \bar{G}_{fi} Y_i + E_{fi}$$

where

$$\bar{\Gamma}_{fi} = CA_K^{i-1} \quad (3.52a)$$

$$\bar{H}_{fi} \triangleq [CA_K^{i-2} B_K \quad \cdots \quad CB_K \quad D] \quad (3.52b)$$

$$\bar{G}_{fi} \triangleq [CA_K^{i-2} K \quad \cdots \quad CK \quad 0] \quad (3.52c)$$

Therefore, through a multi-stage least squares similar to the innovation estimation method, one can estimate $\bar{\Gamma}_f L_z$, \bar{H}_f and \bar{G}_f . $\bar{\Gamma}_f$ can be estimated through a weighted SVD. It is well known that A_K , C , D , B_K , and K can be obtained through the estimates of $\bar{\Gamma}_f$, \bar{H}_f and \bar{G}_f . After that A and B can be backed out through the definition of A_K and B_K .

[Remark 1] The above analysis clearly illustrates the similarity between the innovation estimation method and the whitening filter approach. They all partition the extended state space row-wise and utilize multi-stage least square method to estimate system matrices. The innovation estimation method starts from a state space model in innovations form, while the whitening filter approach is based on a state space model in predictor form.

[Remark 2] There is another implementation of the whitening filter approach [29]. One can estimate the Markov parameters through the high order ARX, and subtracting the effect of future inputs and outputs.

[Remark 3] As pointed out by Chiuso and Picci [9], both approaches require that eigenvalues of A_K lie strictly inside the unit circle. For a finite past horizon, they are biased due to $A_K^p X_{k-p} \neq 0$.

[Remark 4] For finite data the predictor model form is time varying due to a time varying Kalman filter, even though the system is time-invariant. This may complicate the rank condition of $\bar{\Gamma}_f$ and the subsequent extraction of A_k and C from $\bar{\Gamma}_f$. From this point the PARSIM-E is superior to SSARX or the "whitening filter" approach.

[Remark 5] The innovation estimation method uses the process A matrix to form the observability matrix, while the whitening filter approach uses the predictor matrix A_K . For open loop unstable systems the whitening filter approach can be numerically advantageous, as demonstrated in [9]. However, for bounded systems such as stable or integrating systems, this advantage disappears. In the next section we compare these methods using the closed-loop example given in [50] which has one integrating pole and four stable poles. The simulation results seems to favor the innovation estimation method.

3.4.2 Simulation Study

The example in [50] is adopted here for comparison. The model of the plant is given in a state space form:

$$A = \begin{bmatrix} 4.40 & 1 & 0 & 0 & 0 \\ -8.09 & 0 & 1 & 0 & 0 \\ 7.83 & 0 & 0 & 1 & 0 \\ -4.00 & 0 & 0 & 0 & 1 \\ 0.86 & 0 & 0 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0.00098 \\ 0.01299 \\ 0.01859 \\ 0.0033 \\ -0.00002 \end{bmatrix}$$

$$C^T = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, K = \begin{bmatrix} 2.3 \\ -6.64 \\ 7.515 \\ -4.0146 \\ 0.86336 \end{bmatrix}, D = 0$$

The feedback mechanism is

$$u_k = -F(q)y_k + r_k$$

where

$$F(q) = \frac{(0.61q^4 - 2.03q^3 + 2.76q^2 - 1.83q + 0.49)}{q^4 - 2.65q^3 + 3.11q^2 - 1.75q + 0.39} \quad (3.54)$$

and r_k is a zero-mean white noise sequence with standard deviation 1. We take the number of data points $j = 1200$ and generate 100 data sets, each one with the same reference input r_k but with different noise sequence e_k . We choose $f = p = 20$ for innovation estimation methodes, and $f = p = 30$ for whitening filter approaches. In our simulation, we observe that to obtain unbiased estimation the whitening filter approach needs larger f and p than the innovation estimation method. The pole estimation results for the closed-loop experiments are shown in Figs. 3.13, 3.15, 3.17 and 3.19. From the results we can see that all the methods can provide consistent estimates, while the whitening filter approach produce the worst results.

The estimates of the frequency response for the closed-loop simulations are shown in Figs. 3.12, 3.14, 3.16 and 3.18. We can see that the estimated frequency responses from all the methods match well with that of the real system at low frequency, but they all show bias at high frequency.

3.5 Summary

In this chapter, the feasibility of the closed-loop subspace identification is established. It is shown that SIMs are feasible for closed-loop data with roughly the same identifiability requirements as more traditional methods such as PEMs. The key idea is that the consistent identification can be achieved through innovation estimation. The new algorithm is shown to be consistent under certain assumptions. The simulation studies show that the proposed algorithm is consistent under closed-loop conditions, while the traditional SIMs

with CVA weighting fail to provide consistent estimates.

We also analyze another recently proposed closed-loop SIM referred to as the whitening filter approach. The similarity and difference of them are investigated in detail. Both approach partition the extended state space model into block rows and use the information estimated from the first block row further estimate model parameters in the remaining rows. Through this partition the correlation between the process input and innovation due to feedback is decoupled. It turns out that although they are based on different representations of state space models all of them can be implemented through multi-stage least squares. All closed-loop SIMs can be classified into one-step, two-step and multi-stage approaches and each of them seems to have its own advantages.

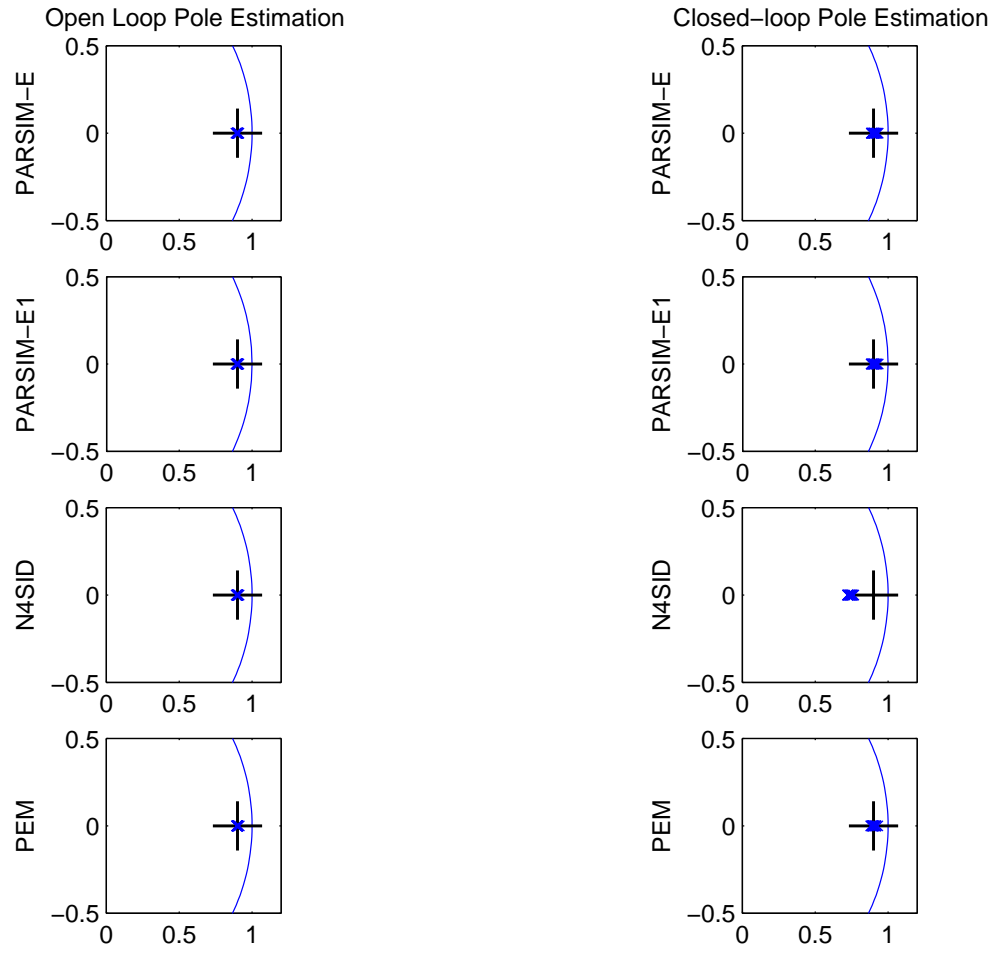


Figure 3.1: Pole estimates for the SISO simulation example

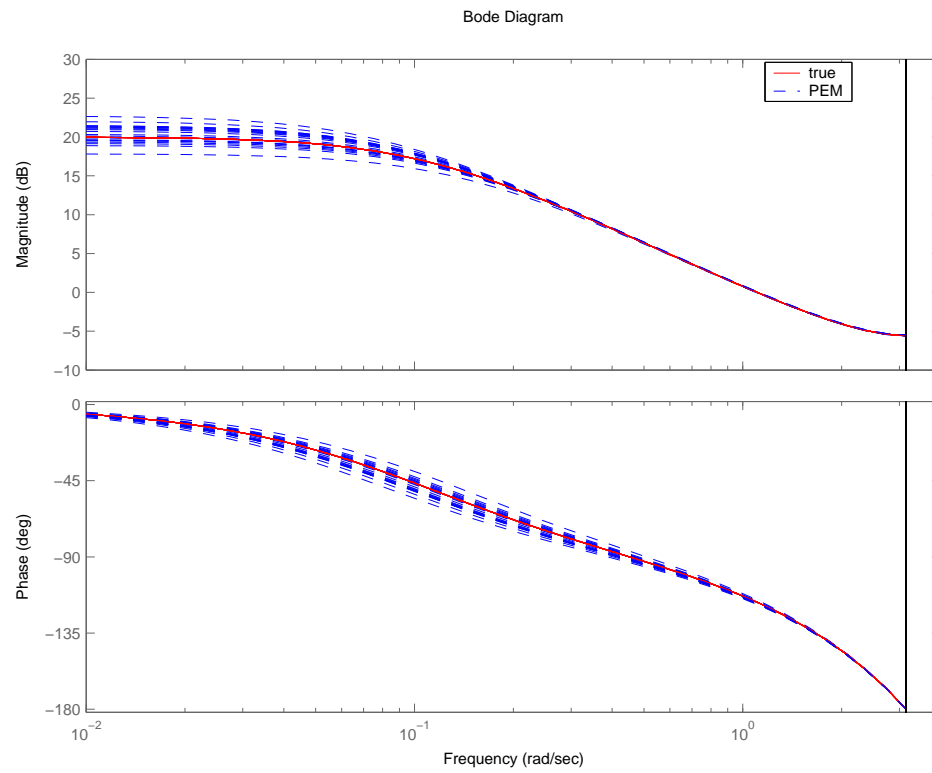


Figure 3.2: The estimates of the frequency response from PEM for SISO closed-loop simulations

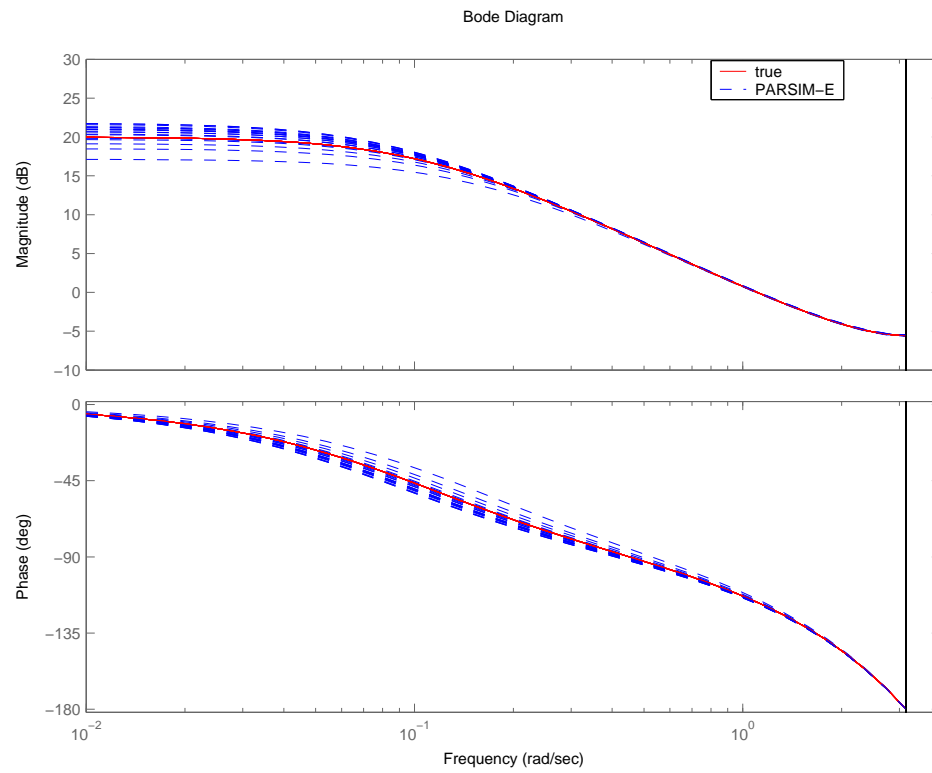


Figure 3.3: The estimates of the frequency response from PARSIM-E SISO closed-loop simulations

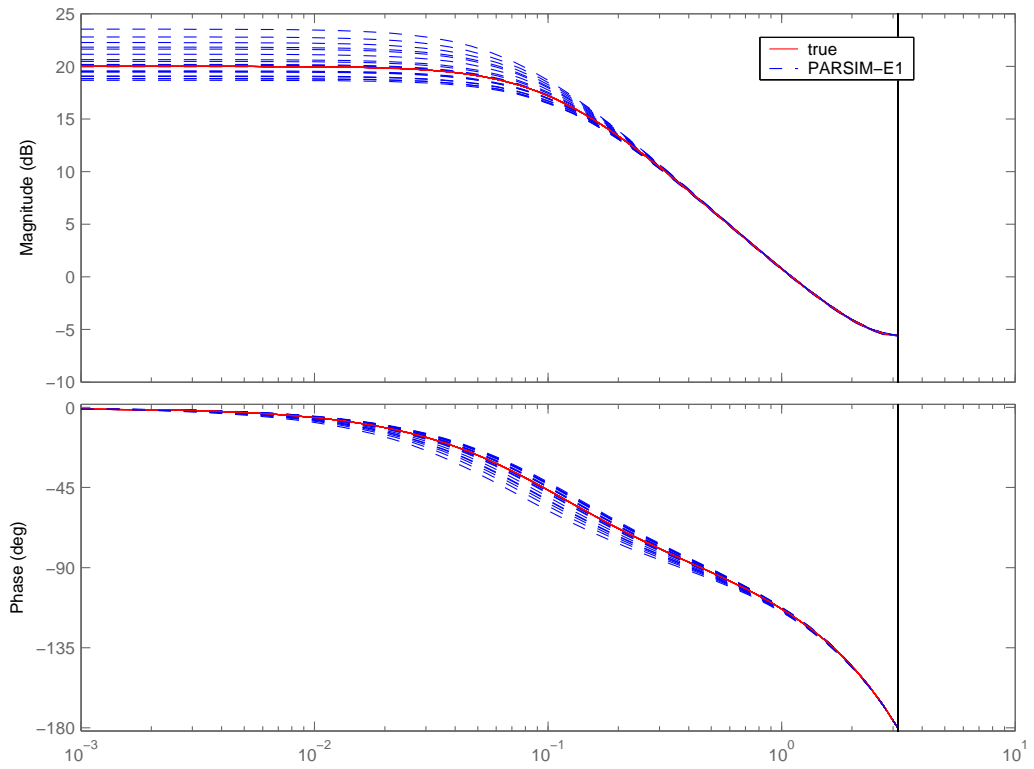


Figure 3.4: The estimates of the frequency response from PARSIM-E1 SISO closed-loop simulations

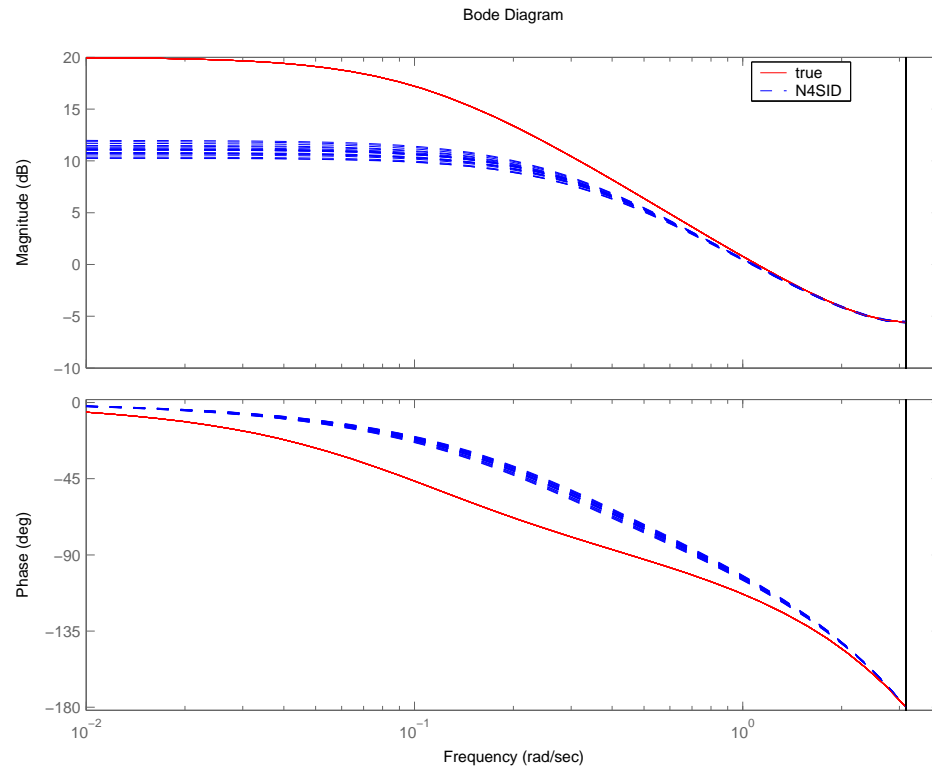


Figure 3.5: The estimates of the frequency response from N4SID for SISO closed-loop simulations

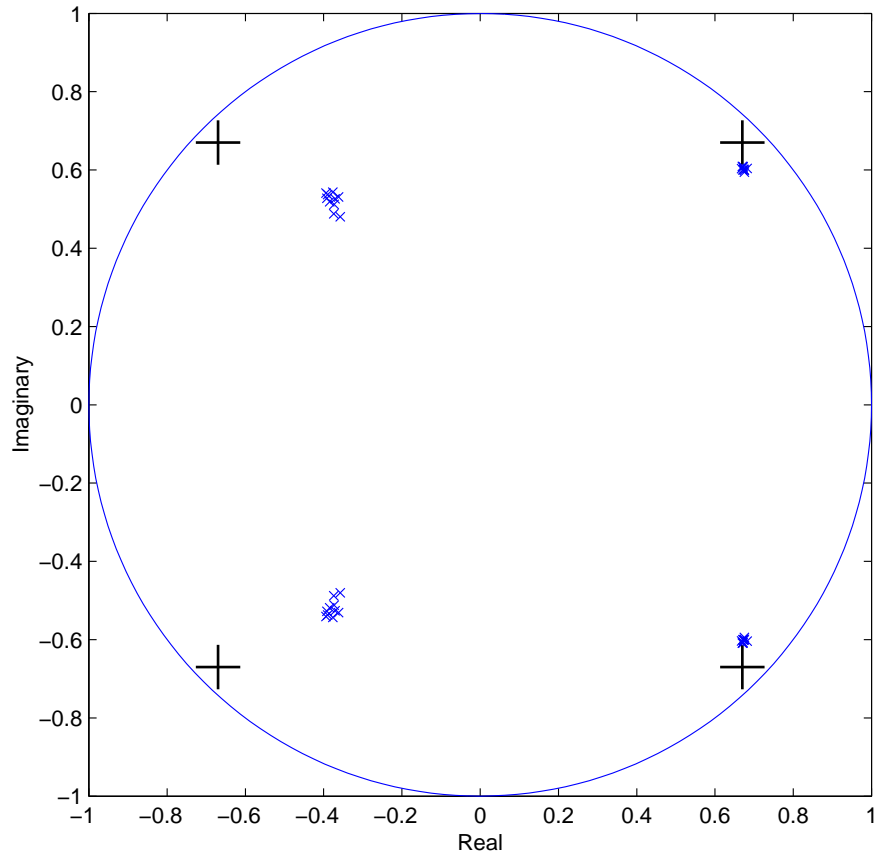


Figure 3.6: The N4SID pole estimation for 10 Monte-Carlo closed-loop simulations: \times estimated pole, $+$ system pole

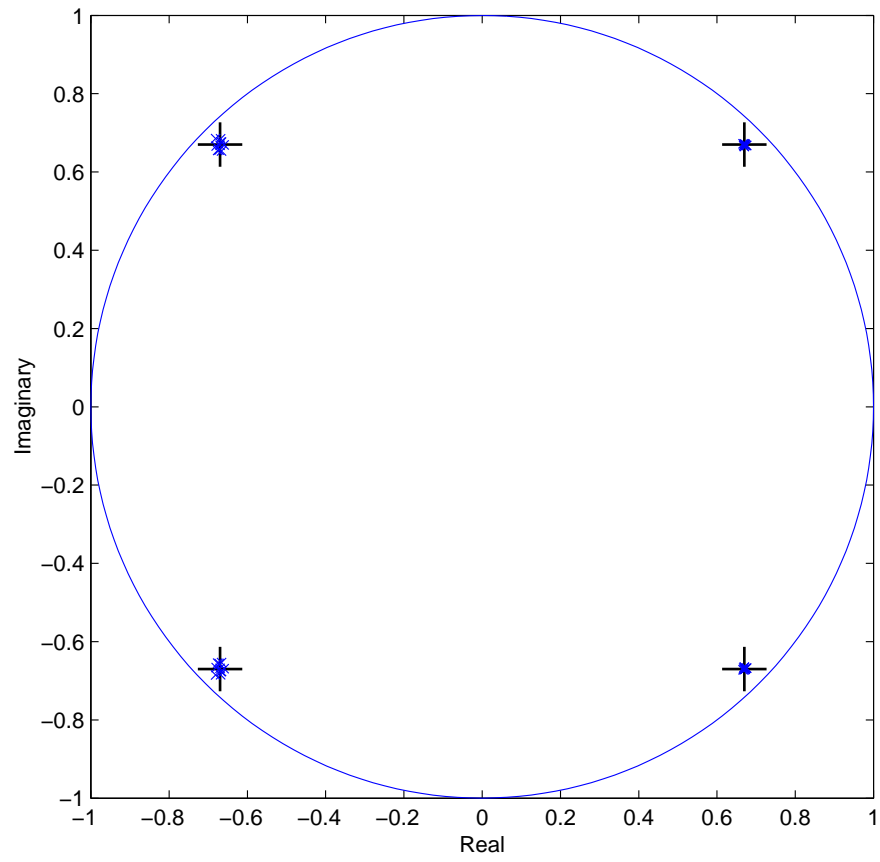


Figure 3.7: The PARSIM-E pole estimation for 10 Monte-Carlo closed-loop simulations: \times estimated pole, $+$ system pole

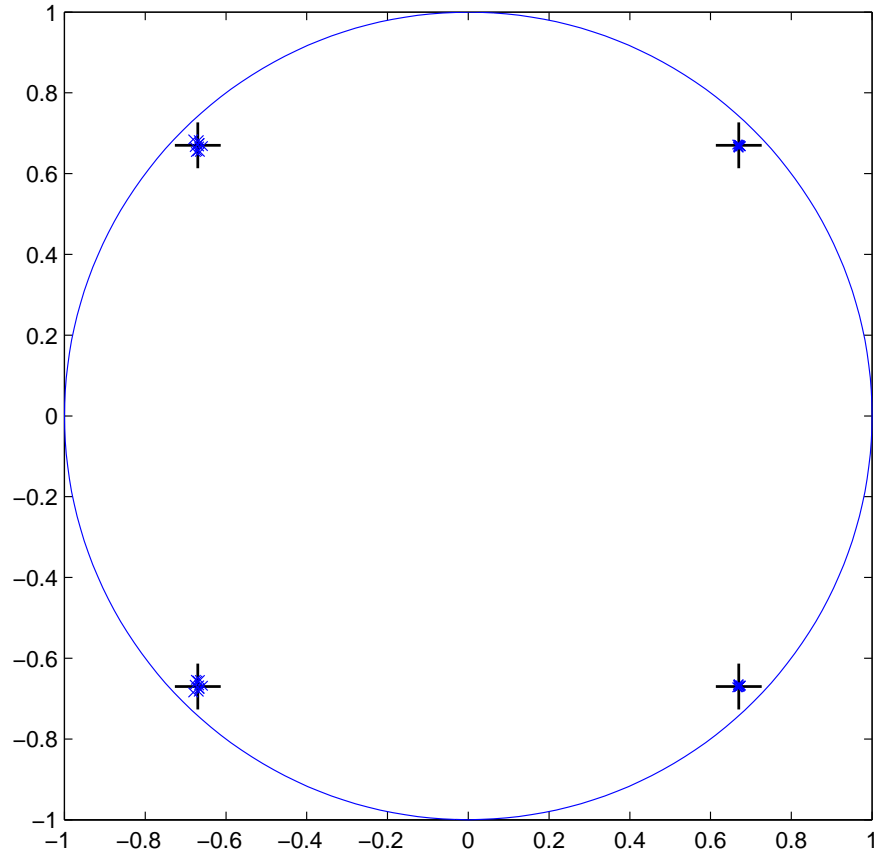


Figure 3.8: The PARSIM-E1 pole estimation for 10 Monte-Carlo closed-loop simulations: \times estimated pole, $+$ system pole

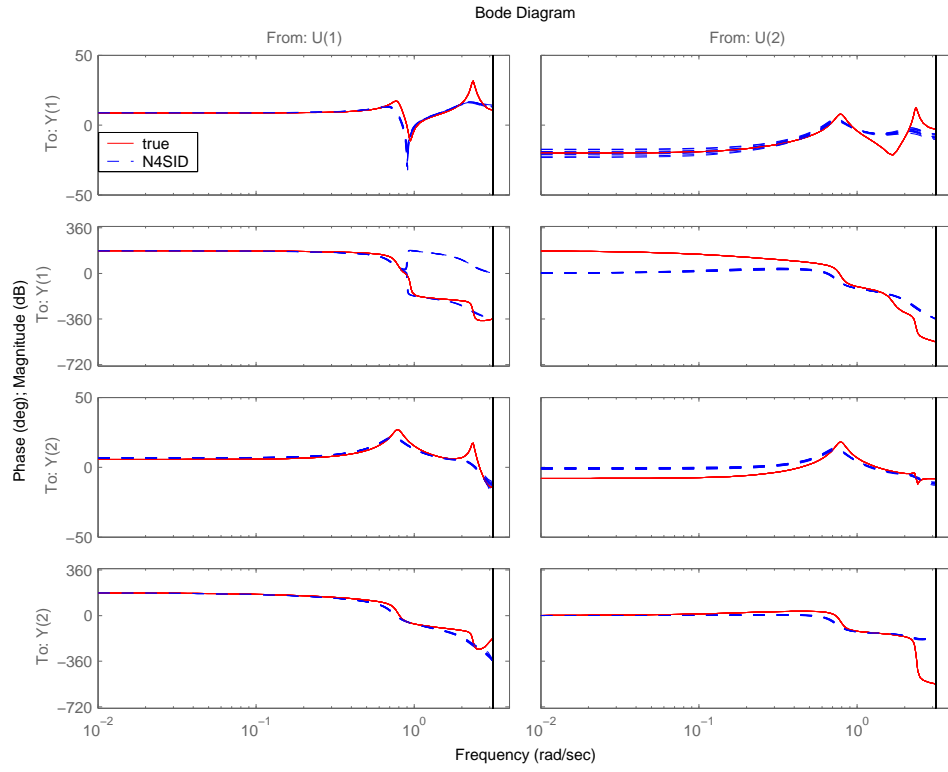


Figure 3.9: The estimates of the frequency response from N4SID for MIMO closed-loop simulations

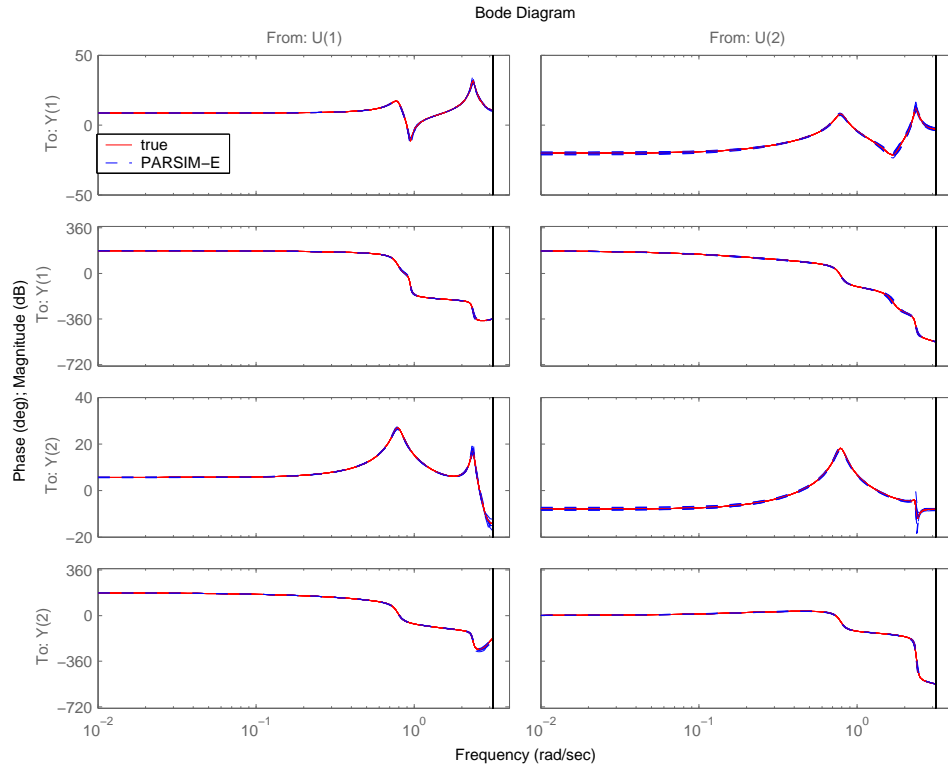


Figure 3.10: The estimates of the frequency response from PARSIM-E for MIMO closed-loop simulations

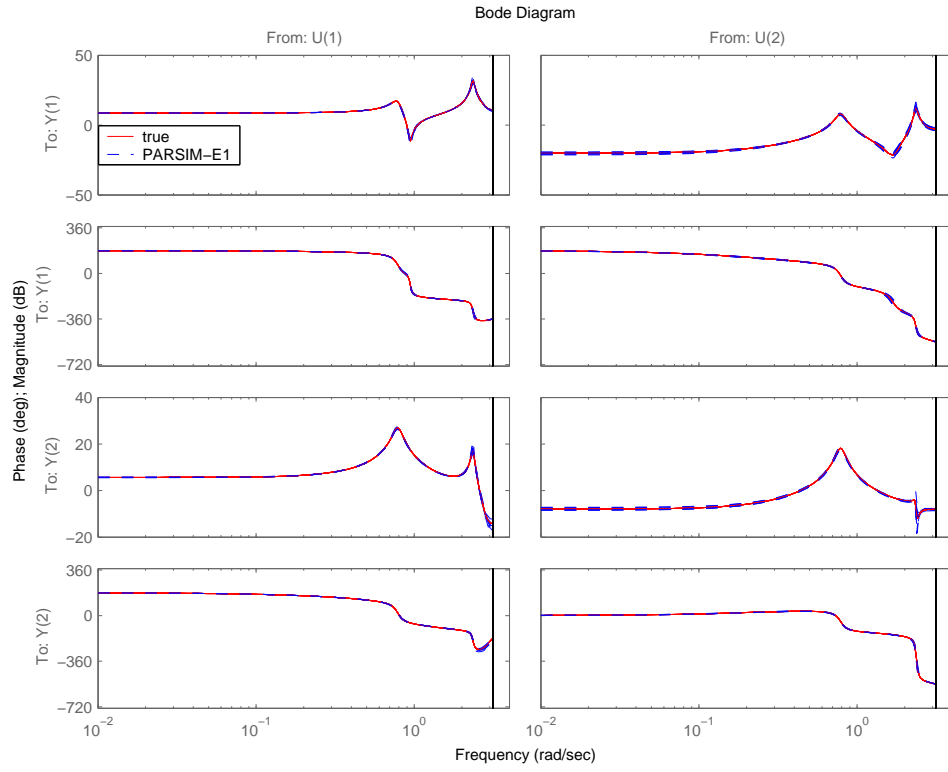


Figure 3.11: The estimates of the frequency response from PARSIM-E1 for MIMO closed-loop simulations

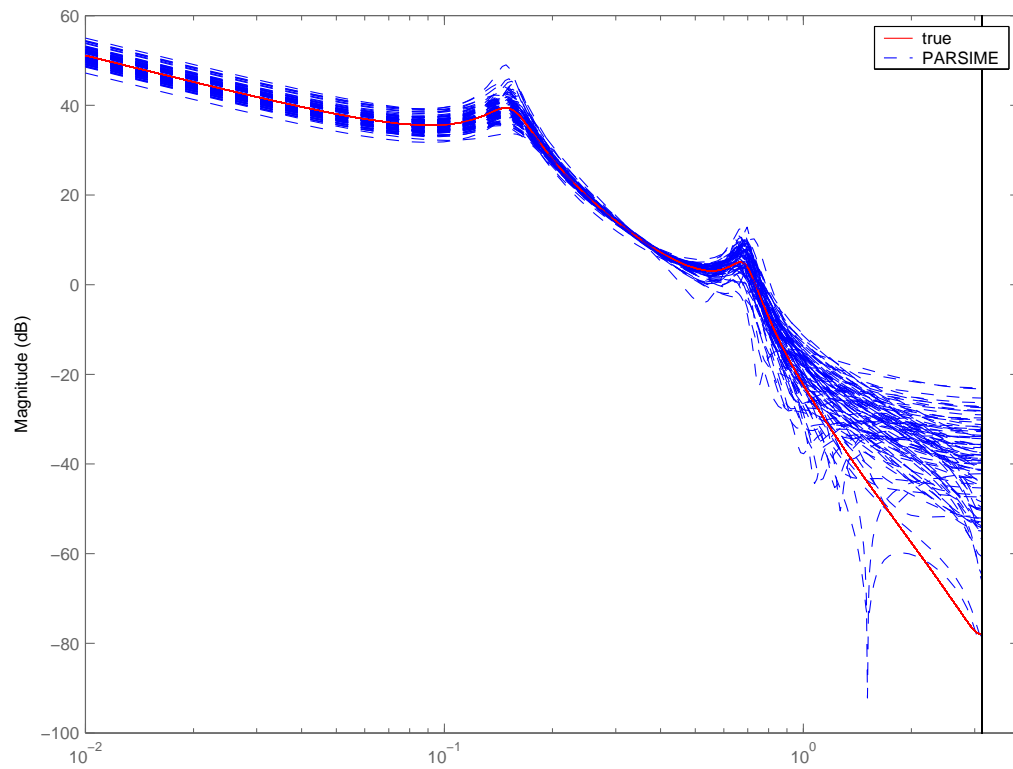


Figure 3.12: The Bode magnitude plot of PARSIM-E for SISO closed-loop simulations.

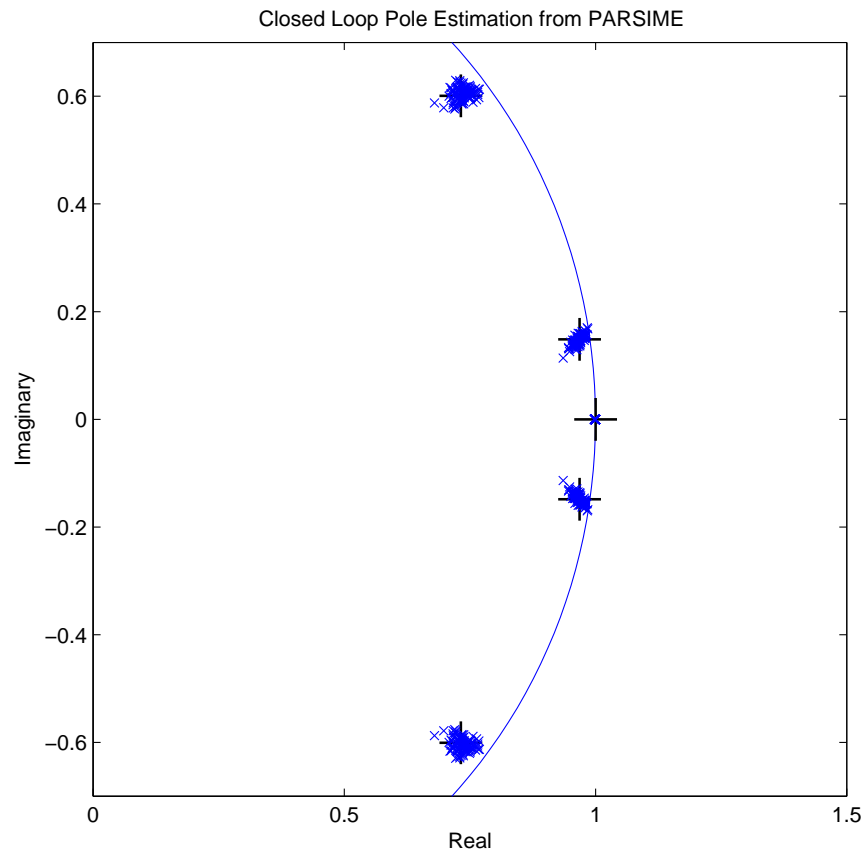


Figure 3.13: The eigenvalues of estimated A matrix from PARSIM-E: \times estimated pole, $+$ system pole.

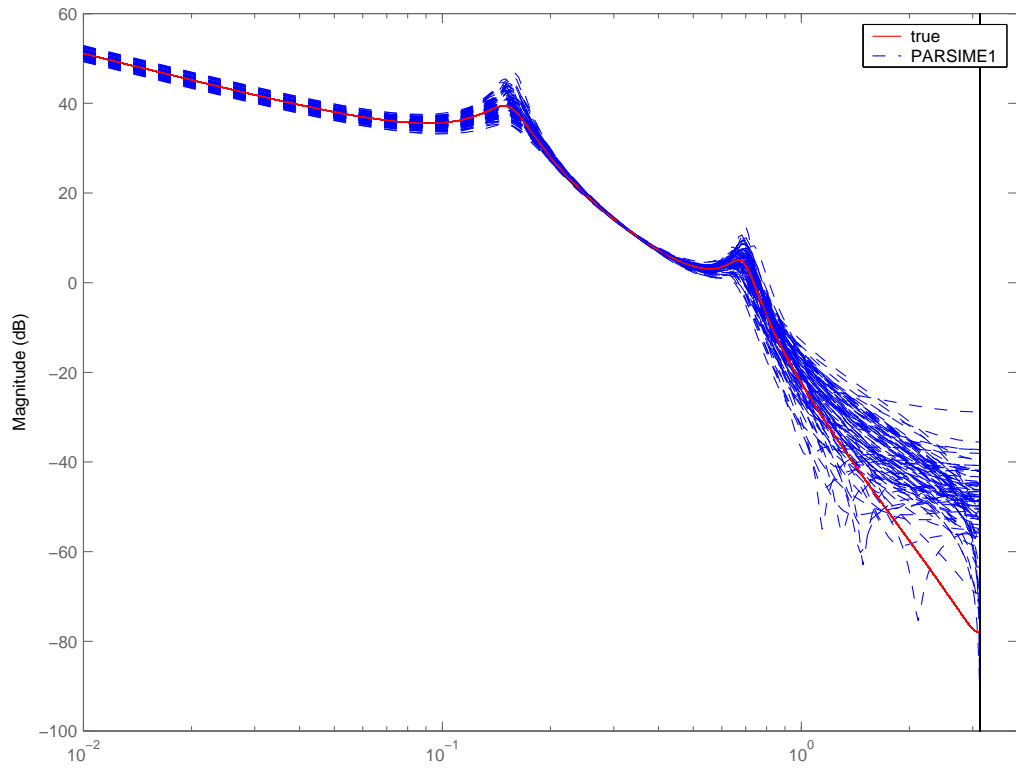


Figure 3.14: The Bode magnitude plot of PARSIM-E1 for SISO closed-loop simulations.

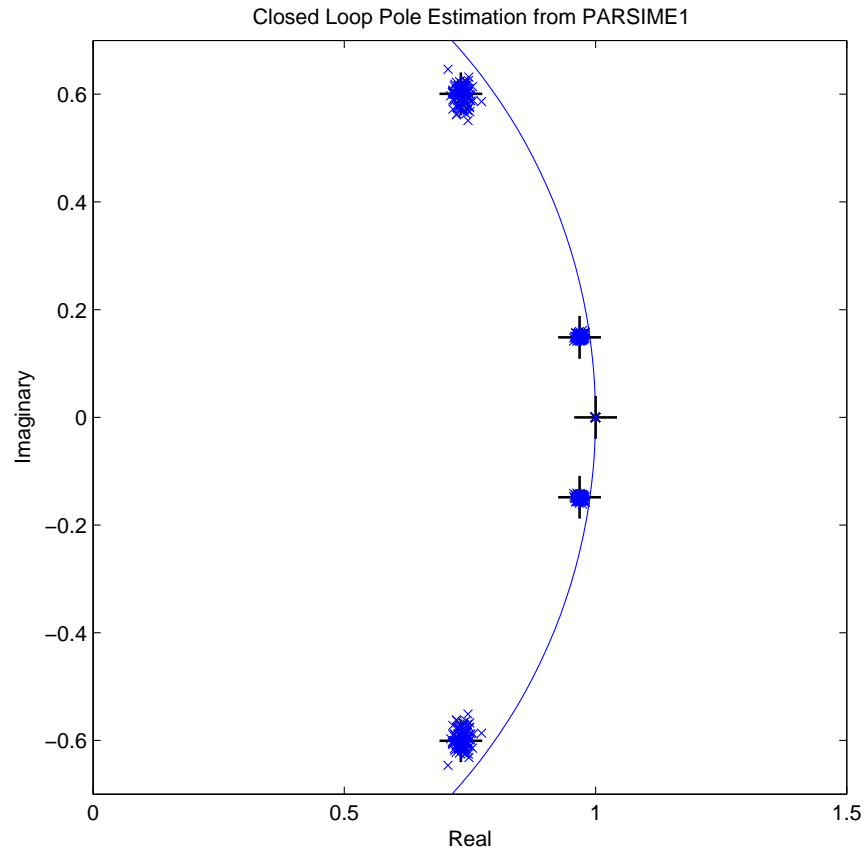


Figure 3.15: The eigenvalues of estimated A matrix from PARSIM-E1: \times estimated pole, $+$ system pole.

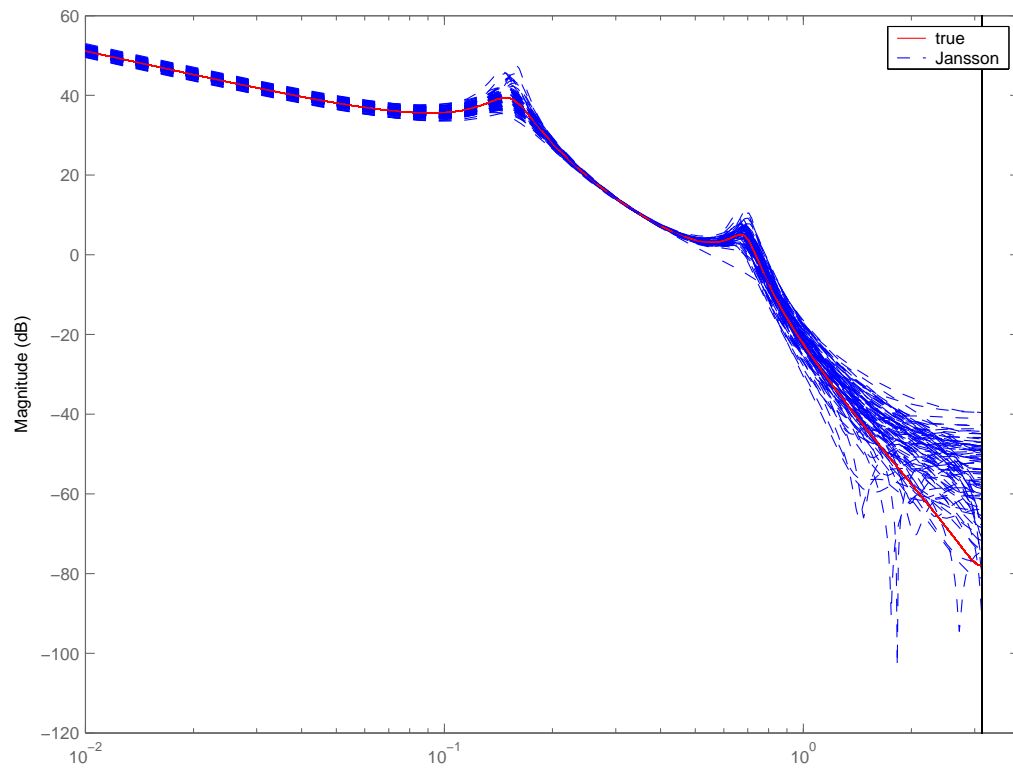


Figure 3.16: The Bode magnitude plot of Jansson's approach for SISO closed-loop simulations.

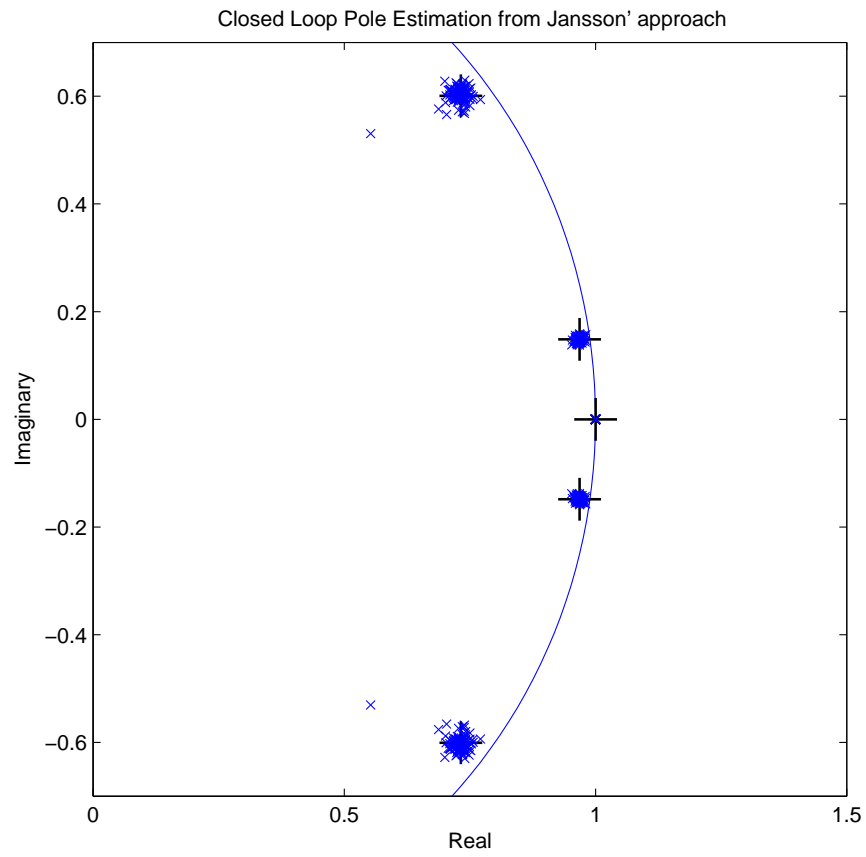


Figure 3.17: The eigenvalues of estimated A matrix from SSARX: \times estimated pole, $+$ system pole.

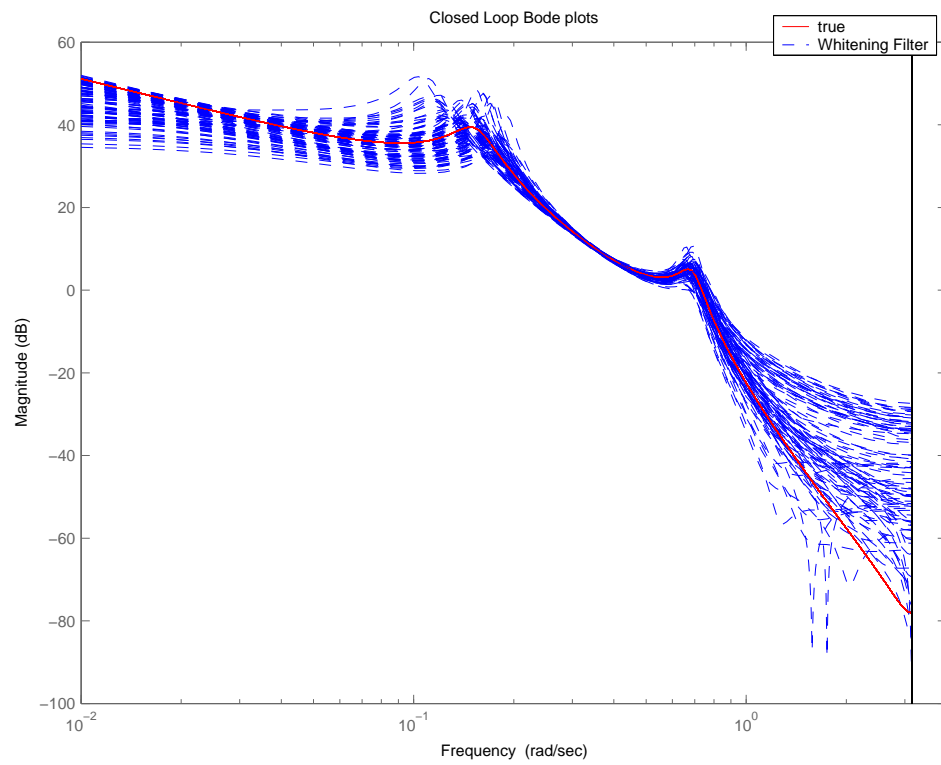


Figure 3.18: The Bode magnitude plot of "whitening filter" approach for SISO closed-loop simulations.

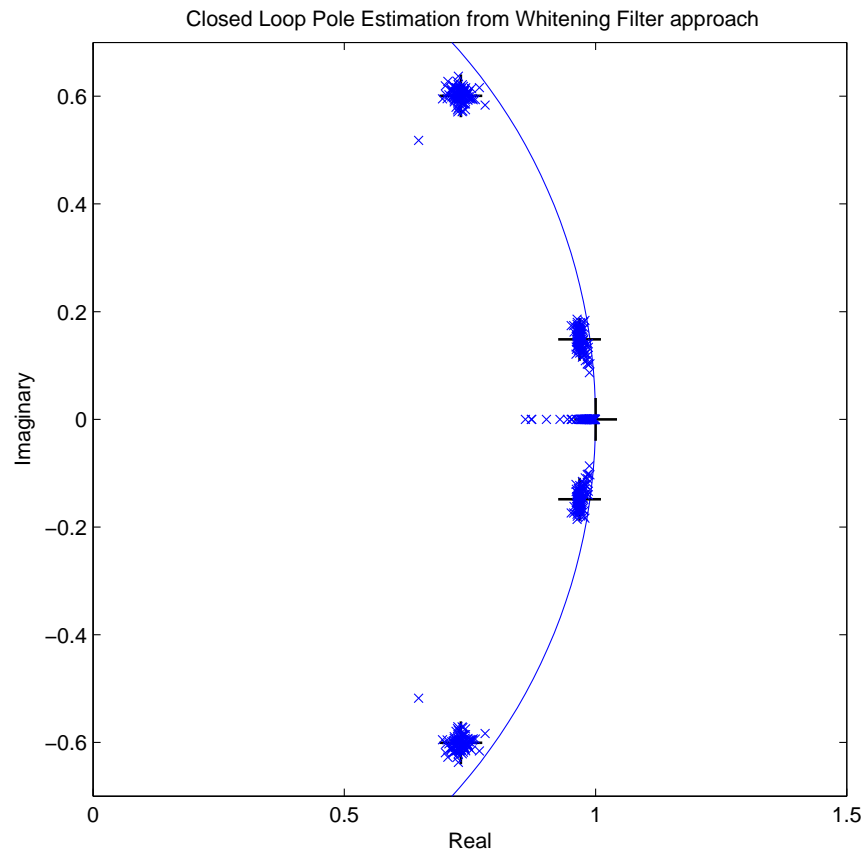


Figure 3.19: The eigenvalues of estimated A matrix from "whitening filter" approach: \times estimated pole, $+$ system pole.

Chapter 4

An Optimal Structured Residual Approach for Unidirectional Faulty Sensor Diagnosis

4.1 Introduction

Measurements in chemical processes, such as temperature, flow rate, and pressure, are subjected not only to random and systematic errors, but also to process and sensor faults. We cannot expect that a set of measurements will exactly obey the physical and chemical principles governing the process when faults occur. In the case of sensor faults, the control system will operate the process based on incorrect information, which will result in product quality variation or even off-specification products. Sensor validation, as an important step in ensuring process integrity and reducing product quality variation, involves detecting and identifying faulty sensors. Once faulty sensors are identified, the validation algorithm should estimate fault magnitudes and replace the faulty measurements with reconstructed values. Therefore the control system can be maintained on-line even though some sensors are faulty.

Much of the early work related to sensor validation falls into the model based approach. The actual behavior of the plant is compared with that predicted on the basis of the mathematical model, which can be obtained by

material and energy balances [13], neural networks [35], principal component analysis (PCA) [16], and subspace identification [56]. Residuals are generated as the difference between actual outputs and those predicted by the model. They are composed of noise, faults, disturbances, and modeling errors. Early efforts to sensor validation can be traced back to data reconciliation in chemical engineering and fault detection in aerospace applications. In chemical engineering applications, usually material and energy balance equations serve as the residuals generating model. Data reconciliation involves optimally adjusting measurements based on model constraints. The work in this area is reviewed thoroughly by Crowe [13]. In aerospace applications, Chow and Willsky [12] have proposed a systematic mechanism to generate analytical redundancy for fault detection and isolation.

While there is an extensive literature in the process monitoring and fault detection area using various approaches, there is relatively little literature dealing with the fault identification issue. Serth and Heenan [60] have proposed a modified iterative measurement test to identify gross errors. They serially eliminated the most suspect measurement to see whether the statistical test violation is removed after the reconciliation. Dunia et al. [16] have proposed a sensor validation index to identify the faulty sensor by reconstructing each sensor in turn based on the PCA model. Yoon and MacGregor [70] have isolated faults through the contribution plot based on the analysis of historical data with PCA and partial least square (PLS). By designing structured residuals that are insensitive to a particular subset of faults, Gertler and

Singer [21] have proposed a structural residuals framework for fault detection and isolation based on the parity equation. Recently, Qin and Li [55, 56] have proposed a structured residuals approach which makes one structured residual insensitive to one subset faults while with maximized sensitivity to other faults.

In this chapter, we propose a new optimal structured residuals approach for unidirectional fault identification. To maximize fault isolation ability, a matrix of optimal structured residuals are designed. Each of them is insensitive to one of faults while being most sensitive to one of remaining ones. The maximum of all structured residuals in each row is then selected as the optimal one for fault isolation. Through this approach, optimal structured residual directions with maximum fault isolation ability are obtained.

The rest of this chapter is organized as follows. In the next section, the fault representation based on the quasi-steady state model is briefly introduced. Our main result, the proposed optimal structured residuals approach is discussed in Section 4.3. Necessary and sufficient conditions for fault isolation are investigated in Section 4.4. A maximum likelihood algorithm for data reconstruction is introduced in Section 4.5. The utility of the proposed approach is illustrated in Section 4.6 using data from an industrial boiler, followed by concluding remarks in Section 4.7.

4.2 Model and fault representation

A quasi-steady state process model can be obtained as follow from first principles or data based approaches,

$$Bx^*(t) = e^*(t) \quad (4.1)$$

where $x^* \in \mathfrak{R}^n$ is a vector of normal sensor measurements, $B \in \mathfrak{R}^{m \times n}$ is the model matrix, and $e^* \in \mathfrak{R}^m$ is the model residual which contains measurement noise, process noise, and possible model mismatch. Under normal conditions the residual vector e^* can be assumed to be zero mean white noise. One should notice that B can be easily obtained from statistical approaches such as principal component analysis and partial least squares.

When sensor faults occur, the measurement can be represented as:

$$x(t) = x^*(t) + \Xi_i f_i(t) \quad (4.2)$$

where $f_i(t) \in \mathfrak{R}^{l_i}$ is a vector of the fault magnitude, $\Xi_i \in \mathfrak{R}^{n \times l_i}$ is a matrix of fault directions, and l_i is the dimension of the fault. The following relation represents the model residual under faulty measurements,

$$e(t) = Bx(t) = Bx^*(t) + B\Xi_i f_i(t) = e^*(t) + B\Xi_i f_i(t) \quad (4.3)$$

The objective of sensor validation is to detect the onset of the fault, identify fault direction matrix, Ξ_i , and estimate the fault magnitude, $f_i(t)$, from faulty measurements.

In order to detect the occurrence of the fault, Qin and Li [55] define a fault detection index as

$$d(t) = e^T(t)R_e^{-1}e(t) \quad (4.4)$$

where $R_e \equiv \mathbf{E}\{e^*(t)e^{*T}(t)\}$ is the covariance matrix of $e^*(t)$ and can be estimated from normal process data. Without sensor faults the fault detection index follows a Chi-square distribution with m degrees of freedom [1].

$$d(t) = e^{*T}(t)R_e^{-1}e^*(t) \sim \chi^2(m) \quad (4.5)$$

As a result, sensor faults can be detected with a certain control limit $d_\alpha = \chi_\alpha^2(m)$, where α is the level of significance. To reduce the effect of transients and noise in measured data, an exponentially weighted moving average (EWMA) filter can be applied to $e(t)$ [55].

4.3 Fault identification with optimal structured residuals

After the detection of faults, it is desirable to identify faulty sensors subsequently. In this section, the structured residual approach is briefly reviewed. The drawback of the traditional approach is illustrated and a new optimal structured residual criterion is introduced in detail.

4.3.1 Review of structured residual approaches

The structured residual approach is introduced by Gerlter and Singer [21]. In their work, the structured residual, $r_i(t)$, is defined as

$$r_i(t) = w_i^T e^*(t) + w_i^T B \Xi_i f_i(t) \quad (4.6)$$

such that

$$w_i^T B \Xi_i = 0$$

where w_i is the structured residual direction, $B \Xi_i$ is the representation of the i^{th} fault direction as defined in (4.3). With this design criterion, the structured residual is insensitive to the i^{th} fault.

As point out by Qin and Li [55], the structured residual approach does not maximize the fault isolation ability. Based on this analysis, they proposed an optimal structured residual approach with maximized sensitivity (SRAMS). The SRAMS direction, w_i , is defined as

$$w_i = \arg \max_{\|w_i\|=1} \|w_i^T [B^\circ \Xi_1, \dots, B^\circ \Xi_n]\|^2 \quad (4.7)$$

such that

$$w_i^T B^\circ \Xi_i = 0$$

where $B^\circ = [b_1^\circ \dots b_n^\circ]$ contains normalized columns of B . With the SRAMS approach, the structured residual defined in (4.6) is insensitive to one of faults but has maximized sensitivity to other faults.

Here we analyze the approach in more detail and point out the possibility of misidentifying faults. For the single fault case, the SRAMS criterion

is equivalent to

$$w_i = \arg \max_{\|w_i\|=1} \|w_i^T B^\circ\|^2 \quad s.t. \quad w_i^T b_i^\circ = 0$$

Geometrically, it makes w_i orthogonal to b_i° and $\sum_{j \neq i} \cos^2 \theta_{ij}$ maximum, where θ_{ij} is the angle between w_i and b_j° . However, it could happen that w_i maximizes the objective function and is orthogonal or nearly orthogonal to another fault, i.e., $b_j^\circ (j \neq i)$. The geometric interpretation is illustrated in Figure 4.1, where w_1 designed to be orthogonal to b_1 based on (4.7) happens to be nearly orthogonal to b_4 . In that case the structure residual (4.6) will be not only insensitive to the 1st fault but also to the 4th fault, which makes them not isolable. To formulate the analysis, we provide a theorem as follows

Theorem 4.3.1. *For a given set of fault direction matrices $\{\Xi_1, \dots, \Xi_n\}$, w_i designed from the SRAMS in (4.7), which is orthogonal to $B\Xi_i$, is also orthogonal to $B\Xi_j (j \neq i)$ if and only if $w'_i = w_i$, where*

$$w'_i = \arg \max_{\|w'_i\|=1} \|w_i'^T [B^\circ \Xi_1, \dots, B^\circ \Xi_{j-1}, B^\circ \Xi_{j+1}, \dots, B^\circ \Xi_n]\|^2 \quad s.t. \quad w_i'^T B^\circ \Xi_i = 0$$

Under this condition, faults Ξ_i and $\Xi_j (j \neq i)$ are not isolable using the SRAMS method.

Proof: See Appendix B.

4.3.2 The design of optimal structured residuals

In this subsection we propose a new optimal structured residuals (OSR) criterion for fault diagnosis. The optimal structured directions can be obtained

as follows:

$$w_i^j = \arg \max_{\|w_i^j\|=1} w_i^{jT} b_j^\circ \quad j = 1, \dots, i-1, i+1, \dots, n \quad (4.8)$$

such that

$$w_i^{jT} b_i^\circ = 0 \quad (4.9)$$

Geometrically, as shown in Figure 4.1, w_i^j is the projection of b_j° onto the plane orthogonal to b_i° . To derive an explicit expression of w_i^j , we choose w_i^j as

$$w_i^j = (I - b_i^\circ b_i^{\circ T}) z_i^j \quad (4.10)$$

such that (4.9) is satisfied, where z_i^j is a vector with appropriate dimension. After introducing a Lagrange multiplier, the objective function can be defined as

$$J = z_i^{jT} (I - b_i^\circ b_i^{\circ T}) b_j^\circ + \frac{1}{2} \lambda (1 - \|(I - b_i^\circ b_i^{\circ T}) z_i^j\|^2)$$

Differentiating the objective function with respect to z_i^j and with the help of (4.10), we can obtain

$$(I - b_i^\circ b_i^{\circ T}) b_j^\circ = \lambda (I - b_i^\circ b_i^{\circ T}) z_i^j = \lambda w_i^j$$

since $\|w_i^j\| = 1$, we have

$$\lambda = \|(I - b_i^\circ b_i^{\circ T}) b_j^\circ\| \quad (4.11)$$

$$w_i^j = \frac{(I - b_i^\circ b_i^{\circ T}) b_j^\circ}{\|(I - b_i^\circ b_i^{\circ T}) b_j^\circ\|} \quad (4.12)$$

Similar to SRAMS [55, 56], the proposed optimal structured residuals approach can be extended to multidimensional faults.

After obtaining the structured direction, the structured residual can be defined as:

$$r_i^j(t) = w_i^{jT} e(t) = w_i^{jT} [e^*(t) + b_k f_k(t)] \quad j = 1, \dots, i-1, i+1, \dots, n \quad (4.13)$$

From the definition, we can see that if $i = k$, the corresponding structured residuals will be $w_i^{jT} e^*(t)$, which is the same as the fault free case, but other residuals $r_i^j(t)$ for $i \neq k$ will be large. We can formulate the statement as

$$r_i^j(t) = \begin{cases} w_i^{jT} e^*(t) & i = k, \\ w_i^{jT} e^*(t) + w_i^{jT} b_k f_k(t) & i \neq k. \end{cases} \quad (4.14)$$

From (4.13), we can obtain a structured residuals matrix as

$$R(t) = \begin{bmatrix} - & r_1^2(t) & r_1^3(t) & \cdots & r_1^n(t) \\ r_2^1(t) & - & r_2^3(t) & \cdots & r_2^n(t) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{n-1}^1(t) & \cdots & r_{n-1}^{n-2}(t) & - & r_{n-1}^n(t) \\ r_n^1(t) & r_n^2(t) & \cdots & r_n^{n-1}(t) & - \end{bmatrix} \quad (4.15)$$

We choose the one with maximum absolute value among each row of $R(t)$ for fault diagnosis.

$$r_i(t) = \max_{j \neq i} |r_i^j(t)| \quad \text{for } i = 1, \dots, n \quad (4.16)$$

If fault k occurs, $r_i^k(t)$ is maximized for the i^{th} row expect for $i = k$. Therefore, eq 4.16 picks the k^{th} column of $R(t)$ which is the largest, that is,

$$r_i(t) = r_i^k(t) = w_i^{kT} [e^*(t) + b_k f_k(t)] \quad (4.17)$$

4.3.3 Fault identification indices

After constructing structured residuals, one needs to design appropriate statistical inference for fault diagnosis. In this subsection, we discuss the deterministic fault and the stochastic fault separately along with identification indices defined by Qin and Li [55].

For deterministic faults, such as bias or drifting, the distribution of $r_i^j(t)$ in (4.14) is

$$r_i^j(t) \sim \begin{cases} N(0, w_i^{jT} R_e w_i^j) & i = k, \\ N(w_i^{jT} b_k \bar{f}_k, w_i^{jT} R_e w_i^j) & i \neq k. \end{cases} \quad (4.18)$$

where \bar{f}_k is the bias of the fault, and R_e is the covariance matrix of normal measurements. After picking up the k^{th} column of $R(t)$ with (4.17), with no fault we have

$$\frac{r_i^2(t)}{w_i^{kT} R_e w_i^k} \sim \chi^2(1),$$

or

$$\frac{r_i^2(t)}{w_i^{kT} R_e w_i^k} \leq \chi_\alpha^2(1)$$

with confidence level α . Therefore, one can define a structured fault identification index as

$$I_{SR}^i(t) = \frac{r_i^2(t)}{w_i^{kT} R_e w_i^k \chi_\alpha^2(1)}. \quad (4.19)$$

Under normal conditions $I_{SR}^i < 1$; if there is a fault in the k^{th} sensor $I_{SR}^i > 1$ except for $i = k$. One can also apply an EWMA filter to structured residuals to obtain a filtered structured residual index (I_{FSR}^i) [55].

For stochastic faults, such as precision degradation, the distribution of $r_i^j(t)$ in (4.14) is

$$r_i^j(t) \sim \begin{cases} N(0, w_i^{jT} R_e w_i^j) & i = k, \\ N(0, w_i^{jT} R_d w_i^j) & i \neq k. \end{cases}$$

where R_d is the covariance matrix of degraded measurements. Assuming that the stochastic fault is independent of $e^*(t)$, the covariance R_d can be derived from (4.17) as follows

$$R_d = R_e + \sigma_{f_k}^2 b_k b_k^T \quad (4.20)$$

where $\sigma_{f_k}^2$ is the variance of the fault. As proposed by Qin and Li [55], one can apply cumulative variance (V_{sum}) index to identify variance change.

$$V_{sum}^i(t) = \sum_{l=t-T}^t (r_i(l) - \hat{\mu}_i)^2$$

such that

$$\hat{\mu}_i = \frac{\sum_{l=t-T}^t r_i(l)}{T+1}$$

where T is the moving window size. After picking up the k^{th} column of $R(t)$ with (4.17), with no fault we have,

$$\frac{V_{sum}^i(t)}{w_i^{kT} R_e w_i^k} \sim \chi^2(T)$$

Therefore one can define a V_{sum} index as

$$I_{V_{sum}}^i(t) = \frac{V_{sum}^i(t)}{w_i^{kT} R_e w_i^k \chi_\alpha^2(T)} \quad (4.21)$$

If $I_{V_{sum}}^i > 1$ for all i but $i = k$, the k^{th} sensor has a variance fault.

After detecting the occurrence of faults, a faulty sensor can be identified as follows

1. Based on the process model matrix B , calculate optimal structured directions w_i^j off-line.
2. From the process residual $e(t)$, calculate optimal structured residuals $r_i(t)$, and corresponding identification indices.
3. If the i^{th} index is less one while other indices are greater than one, the i^{th} sensor is faulty.

4.4 Fault isolability

The fault isolability issue with structured residuals is discussed by Gertler [20] in the transfer matrix form. Here we investigate the issue and discuss how to deal with it for the OSR method.

As we can see, if two columns of B , i.e., b_i and b_j , are exactly collinear, the i^{th} fault and the j^{th} fault are not isolable. Therefore, the necessary condition for the i^{th} and the j^{th} faults to be isolable is that b_i and b_j are not collinear. When two fault directions are nearly collinear, they are difficult to isolate, unless the fault magnitude is sufficiently large.

If a variable has very small coefficients in (4.1), i.e., b_k has a very small norm, the k^{th} sensor faults is difficult to detect or identify. This can happen to models derived from first principles or from data based methods such as PCA [54]. For a PCA model $B = \tilde{P}^T$ [55], where \tilde{P} is the residual loading matrix corresponding to the smallest singular values. The matrix \tilde{P}^T can have a near zero column if the corresponding principal loading matrix P^T has a column

with a norm close to one. In this case the sensor fault corresponding to this column is hard to detect or identify.

To measure the collinearity between two columns b_i and b_k , we denote the angle between them as β_{ik} , and the angel between b_k and w_i^k as θ_{ik} . It is easy to show that

$$\theta_{ik} = \begin{cases} 90^\circ - \beta_{ik} & \beta_{ik} \leq 90^\circ, \\ \beta_{ik} - 90^\circ & \beta_{ik} > 90^\circ. \end{cases}$$

since $w_i^k \perp b_i$. If b_i and b_k are nearly collinear, β_{ik} is close to 0° , making θ_{ik} close to 90° . For a bias type of fault the mean shift due to the fault is derived from (4.17) as

$$E[r_i(t)] = w_i^{kT} b_k \bar{f}_k = \|b_k\| \bar{f}_k \cos \theta_{ik}$$

Therefore, if θ_{ik} is close to 90° , the impact of the fault can hardly be observed.

We can discuss the sufficient condition for fault isolability, which is how large the fault magnitude should be to guarantee fault isolability. When fault k occurs, from (4.17) and (4.18) we have

$$r_i(t) = r_i^k(t) \sim \begin{cases} N(0, w_i^{kT} R_e w_i^k) & i = k, \\ N(w_i^{kT} b_k \bar{f}_k, w_i^{kT} R_e w_i^k) & i \neq k. \end{cases} \quad (4.22)$$

Therefore, we can obtain [32]

$$Y_i(t) = \frac{r_i^2(t)}{w_i^{kT} R_e w_i^k} \sim \begin{cases} \chi^2(1) & i = k, \\ \chi^2(1, \lambda_i) & i \neq k. \end{cases} \quad (4.23)$$

where $\chi^2(1, \lambda_i)$ is the the non-central χ^2 with one degree of freedom and non-centrality

$$\lambda_i = \frac{(w_i^{kT} b_k \bar{f}_k)^2}{w_i^{kT} R_e w_i^k}$$

To guarantee that the i^{th} fault is isolated from the k^{th} fault with some confidence level β , we must have

$$P\{I_{SR}^i(t) = \frac{Y_i(t)}{\chi_\alpha^2(1)} > 1\} = 1 - P[Y_i(t) \leq \chi_\alpha^2(1)] = \beta$$

or $P[Y_i(t) \leq \chi_\alpha^2(1)] = 1 - \beta$. From (4.23) we know that

$$P[Y_i \leq \chi_{1-\beta}^2] = 1 - \beta$$

Therefore, if we find a critical value, λ_{ic} , such that

$$\chi_{1-\beta}^2(1, \lambda_{ic}) = \chi_\alpha^2(1),$$

then $I_{SR}^i(t) > 1$ with confidence level β for $\lambda_i > \lambda_{ic}$.

With the help of (4.23), we can derive the sufficient condition for deterministic fault isolability as

$$|\bar{f}_k| > \frac{[w_i^{kT} R_e w_i^k \lambda_{ic}]^{\frac{1}{2}}}{w_i^{kT} b_k} = \frac{[w_i^{kT} R_e w_i^k \lambda_{ic}]^{\frac{1}{2}}}{\|b_k\| \cos \theta_{ik}}, \quad (4.24)$$

From the above equation, we can see that two important factors are $\|b_k\|$ and θ_{ik} .

The sufficient isolability condition for the variance type of fault can be derived as follows. When fault k occurs and $i = k$ we know that

$$P\left\{\frac{V_{sum}^i(t)}{w_i^{kT} R_e w_i^k} \leq \chi_\alpha^2(T)\right\} = \alpha$$

or

$$P\{V_{sum}^i \leq w_i^{kT} R_e w_i^k \chi_\alpha^2(T)\} = \alpha$$

For another sensor i ($i \neq k$),

$$\frac{V_{sum}^i(t)}{w_i^{kT} R_d w_i^k} \sim \chi^2(T) \quad \text{for all } i \neq k$$

Therefore

$$P\left\{\frac{V_{sum}^i(t)}{w_i^{kT} R_d w_i^k} \geq \chi_{1-\beta}^2(T)\right\} = \beta$$

or

$$P\left\{I_{V_{sum}}^i(t) = \frac{V_{sum}^i(t)}{w_i^{kT} R_e w_i^k \chi_\alpha^2(T)} \geq \frac{w_i^{kT} R_d w_i^k \chi_{1-\beta}^2(T)}{w_i^{kT} R_e w_i^k \chi_\alpha^2(T)}\right\} = \beta, \quad \text{for } i \neq k$$

To guarantee the $I_{V_{sum}}^i(t) > 1$ with some confidence level β , we require

$$w_i^{kT} R_d w_i^k \chi_{1-\beta}^2(T) > w_i^{kT} R_e w_i^k \chi_\alpha^2(T)$$

Using the result of (4.20) we obtain

$$\begin{aligned} \sigma_{f_k}^2 &> \frac{w_i^{kT} R_e w_i^k [\chi_\alpha^2(T) - \chi_{1-\beta}^2(T)]}{w_i^{kT} b_k b_k^T w_i^k \chi_{1-\beta}^2(T)} \\ &= \frac{w_i^{kT} R_e w_i^k [\chi_\alpha^2(T) - \chi_{1-\beta}^2(T)]}{\chi_{1-\beta}^2(T) \|b_k\|^2 \cos^2 \theta_{ik}} \end{aligned} \quad (4.25)$$

Again $\|b_k\|$ and θ_{ik} play an important role in the isolation of the k^{th} fault from the i^{th} fault.

From a practical point of view, in order to make B matrix satisfy fault isolability conditions, once B matrix is obtained, it is always a good practice to check the angles between every two columns and the norm of the columns. It will provide information about the degree of difficulty for fault isolation.

4.5 Reconstruction of normal measurements

After identifying the faulty sensor, it is important to determine the necessary adjustment and to bring the measurement back to normal. Dunia and Qin [15] proposed a least-squares solution to estimate the fault magnitude and investigated the reconstructability issue for the model obtained from PCA. In this section, the maximum likelihood estimation of fault magnitude is proposed for the general linear model.

If there are sensor faults, the mean of actual residual $e(t)$ will increase according to (4.3) as

$$e(t) = e^*(t) + B\Xi_i f_i(t) \sim N(B\Xi_i f_i(t), R_e) \quad (4.26)$$

where R_e is the covariance matrix of $e^*(t)$. Since the fault direction matrix, Ξ_i , has been identified, the fault magnitude can be obtained through maximum likelihood estimation

$$\hat{f}_i(t) = \arg \min_{f_i} \|Q[B\Xi_i f_i(t) - e(t)]\|^2 \quad (4.27)$$

where $Q^T Q = R_e^{-1}$ is the Choleskey factorization of R_e^{-1} . It is straightforward that if $B\Xi_i$ is full column rank, the estimate of the fault magnitude is the maximum likelihood estimation

$$\hat{f}_i(t) = (\Xi_i^T B^T R_e^{-1} B \Xi_i)^{-1} \Xi_i^T B^T R_e^{-1} e(t) \quad (4.28)$$

If $B\Xi_i$ is not full column rank, the estimate of the fault magnitude is the minimum 2-norm solution [22] of (4.27)

$$\hat{f}_i(t) = V_1 \Sigma_r^{-1} U_1^T Q e(t) \quad (4.29)$$

where

$$QB\Xi_i = [U_1 \ U_2] \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}$$

is the singular value decomposition of $QB\Xi_i$. According to (4.2), the reconstructed measurements are

$$\hat{x}^*(t) = x(t) - \Xi_i \hat{f}_i(t) \quad (4.30)$$

4.6 An industrial boiler case study

The proposed fault identification and reconstruction scheme are applied to an industrial boiler described in Qin and Li [55] with seven measured variables. Over 630 data points are collected at a five-minute sampling interval. The seven variables considered are list in Table 4.1. The data are scaled to zero mean and unit variance, and then the process model is built using PCA. The number of principal components is determined based on the best reconstruction criterion [54] and the identified process model, B , is derived as in [55].

Variable name	Sensor no.	Minimum	Maximum	Mean	Std
Air flow	1	215.88	415.74	314.98	44.84
Fuel flow	2	10.48	20.13	16.43	2.11
Steam flow	3	150.98	300.62	244.12	33.14
Economizer temperature	4	622.66	737.81	699.53	23.67
Stack pressure	5	2.02	10.79	7.11	1.93
Wind-box pressure	6	2.69	11.30	7.52	2.00
Feed water flow	7	172.97	308.18	253.58	31.14

Table 4.1: The process variables for the boiler process

$$B = \begin{bmatrix} -0.1466 & -0.0950 & -0.1198 & -0.1282 & -0.3028 & -0.1012 & 0.9147 \\ -0.1017 & -0.0788 & 0.1813 & 0.8058 & -0.3142 & -0.4481 & -0.0414 \\ 0.0967 & 0.6896 & 0.3149 & -0.2982 & -0.5361 & -0.1660 & -0.1093 \\ -0.1522 & 0.1498 & 0.2087 & -0.2053 & 0.5983 & -0.7062 & 0.1097 \\ -0.7627 & 0.4782 & -0.2731 & 0.1977 & 0.1347 & 0.2403 & -0.0094 \\ -0.4570 & -0.3376 & 0.7684 & -0.1486 & -0.0553 & 0.2473 & -0.0194 \end{bmatrix} \quad (4.31)$$

In this work, two kinds of faults are simulated: bias and precision degradation, which can be represented as

$$f_i(t) = C_1 \quad (4.32)$$

$$f_i(t) \sim N(0, C_2) \quad (4.33)$$

where C_1 and C_2 are constants for different faults, and f_i is the fault magnitude. Bias and precision degradation faults represent changes in mean and variance, which are suitable for EWMA filters and cumulative variance indices. Drift and complete failure considered in Qin and Li [55] affect both the mean and variance, which can be identified by both EWMA and cumulative variance indices. For this reason we consider bias and precision degradation only in this paper. In order to compare the fault identification ability of different approaches, single sensor faults with different fault magnitudes are introduced at the 200th sample (4.2). As our purpose is not to compare fault detection ability of different approaches, we assume that the fault is detected at the 220th sample.

To compare the fault identification ability of OSR and SRAMS, two kinds of fault identification indices are investigated, the filtered structured

	Faulty Sensor						
	1	2	3	4	5	6	7
	O/S	O/S	O/S	O/S	O/S	O/S	O/S
0.5 %	-/-	-/-	-/-	-/-	-/-	-/-	-/-
1.5 %	√/-	-/-	-/-	√/-	-/-	-/-	-/-
2.5 %	√/-	√/-	√/-	√/-	-/-	-/-	-/-
3.5 %	√/-	√/-	√/-	√/-	-/-	-/-	-/-
4.5 %	√/√	√/-	√/-	√/-	-/-	-/-	√/-
7.5 %	√/√	√/-	√/√	√/-	√/-	-/-	√/-
8.5 %	√/√	√/-	√/√	√/-	√/-	√/-	√/-

Table 4.2: Fault identification results for the single biased sensor with I_{FSR} through OSR (O) and SRAMS (S) approaches. $\sqrt{}$ means the faulty sensor can be correctly identified, while $-$ means the faulty sensor cannot be identified.

residuals index (I_{FSR}) with EWMA filter coefficient 0.65, and the cumulative variance index ($I_{V_{sum}}$) with a moving window size of 20 are used in this work.

They are applied to identify bias and precision degradation, respectively.

Faulty Sensor						
1	2	3	4	5	6	7
2.7%	4.0%	3.5%	1.6%	11.8%	10.6%	7.6%

Table 4.3: The minimum bias fault required by fault isolability condition with $\alpha = 0.95$ and $\beta = 0.90$, which is calculated as the percentage of the mean of the corresponding sensor.

With the process model matrix defined in (4.31), the optimal structured residual matrix can be calculated as

$$R(t) = \begin{bmatrix} - & r_1^2(t) & r_1^3(t) & r_1^4(t) & r_1^5(t) & r_1^6(t) & r_1^7(t) \\ r_2^1(t) & - & r_2^3(t) & r_2^4(t) & r_2^5(t) & r_2^6(t) & r_2^7(t) \\ r_3^1(t) & r_3^2(t) & - & r_3^4(t) & r_3^5(t) & r_3^6(t) & r_3^7(t) \\ r_4^1(t) & r_4^2(t) & r_4^3(t) & - & r_4^5(t) & r_4^6(t) & r_4^7(t) \\ r_5^1(t) & r_5^2(t) & r_5^3(t) & r_5^4(t) & - & r_5^6(t) & r_5^7(t) \\ r_6^1(t) & r_6^2(t) & r_6^3(t) & r_6^4(t) & r_6^5(t) & - & r_6^7(t) \\ r_7^1(t) & r_7^2(t) & r_7^3(t) & r_7^4(t) & r_7^5(t) & r_7^6(t) & - \end{bmatrix}$$

where

$$r_i^j(t) = w_i^j e(t)$$

and

$$w_i^j = \frac{(I - b_i^o b_i^{oT}) b_j^o}{\|(I - b_i^o b_i^{oT}) b_j^o\|} \quad i = 1 \dots 7; j = 1, \dots, i-1, i+1, \dots, 7$$

We then pick up the maximum of each row as $r_i(t)$ according to (4.16) and the corresponding fault identification indices, the single faulty sensor can be identified as discussed at the end of Section 4.3.

	Faulty Sensor						
	1	2	3	4	5	6	7
	O/S	O/S	O/S	O/S	O/S	O/S	O/S
5 %	-/-	-/-	-/-	-/-	-/-	-/-	-/-
10 %	✓/✓	-/-	✓/-	-/-	-/-	-/-	-/-
15 %	✓/✓	✓/-	✓/-	-/-	-/-	-/-	-/-
20 %	✓/✓	✓/-	✓/✓	-/-	✓/-	✓/-	-/-
25 %	✓/✓	✓/-	✓/✓	✓/-	✓/-	✓/-	-/-
30 %	✓/✓	✓/-	✓/✓	✓/-	✓/-	✓/-	✓/-

Table 4.4: Fault identification results for the single precision degraded sensor with $I_{V_{sum}}$ through OSR(O) and SRAMS (S) approaches. ✓ means the faulty sensor can be correctly identified, while – means the faulty sensor cannot be identified.

Faulty Sensor						
1	2	3	4	5	6	7
11.8%	18.0%	13.6%	25.1%	22.5%	19.3%	36.2%

Table 4.5: The minimum precision degradation fault required by fault isolability condition with $\alpha = 0.95$ and $\beta = 0.90$, which is calculated as the percentage of the standard deviation of the corresponding sensor.

The fault identification results for a single biased sensor fault with OSR and SRAMS are shown in Table 4.2. Fault magnitudes are calculated as the percentage of the mean of corresponding sensor measurements. As we can see that OSR can identify biased sensors with smaller fault magnitude than SRAMS does. Compared with the minimum bias required for fault isolability list in Table 4.3, we can see that the simulation results match approximately with the theoretical calculation from (4.24).

Table 4.4 illustrates identification results for the precision degraded sensor with OSR and SRAMS, respectively. Fault magnitudes in this case are the percentage of stand deviation of the corresponding sensor measurements. We can conclude that the OSR outperforms SRAMS on identifying the precision degraded sensor as well. Comparing with the minimum sensor precision degradation required by fault isolability condition (4.25) listed in Table 4.5, we can see that the simulation results match approximately with the theoretical calculation.

In our simulation, we notice that even as the fault magnitude increases, SRAMS cannot identify the 7th sensor fault. Here we investigate the problem

in more detail. After calculating the following matrix

$$\{\theta_{ij}\} = \begin{bmatrix} 90.0^\circ & 113.3^\circ & 18.4^\circ & 95.5^\circ & 98.2^\circ & 106.6^\circ & 91.9^\circ \\ 34.6^\circ & 90.0^\circ & 134.3^\circ & 91.4^\circ & 92.0^\circ & 93.3^\circ & 90.5^\circ \\ 17.4^\circ & 112.1^\circ & 90.0^\circ & 96.1^\circ & 98.9^\circ & 106.6^\circ & 92.2^\circ \\ 147.4^\circ & 85.9^\circ & 47.8^\circ & 90.0^\circ & 88.1^\circ & 86.7^\circ & 89.5^\circ \\ 146.9^\circ & 85.9^\circ & 47.4^\circ & 88.6^\circ & 90.0^\circ & 86.7^\circ & 89.5^\circ \\ 145.9^\circ & 85.9^\circ & 46.4^\circ & 88.6^\circ & 88.0^\circ & 90.0^\circ & 89.5^\circ \\ 148.0^\circ & 85.9^\circ & 48.4^\circ & 88.6^\circ & 88.1^\circ & 86.7^\circ & 90.0^\circ \end{bmatrix}$$

where θ_{ij} is the angle between w_i and b_j° , and w_i is the i^{th} structured direction [55]. We can see that b_7° is almost orthogonal to every structured residual direction. Therefore when the 7th sensor is faulty, all structured residuals are statistically insignificant, which makes the 7th fault not isolable from others.

As illustrated in (4.28) in Section 4, the fault estimation algorithm is independent of the fault types. To compare the performance of different reconstruction algorithms, we define the reconstructed mean square error (MSE) as

$$MSE = \frac{1}{N - t_f + 1} \sum_{t=t_f}^N [\hat{x}_i^*(t) - x_i^*(t)]^2$$

where $\hat{x}_i^*(t)$ is the reconstructed normal measurement and x_i^* is the normal measurement for the i^{th} faulty sensor. N is the sample size, and t_f is the fault detection time, respectively.

	Faulty Sensor						
	1	2	3	4	5	6	7
ML estimates	0.0017	0.0079	0.0028	0.0161	0.0127	0.0065	0.1002
LS estimates	0.0053	0.0168	0.0068	0.0279	0.0168	0.0137	0.1006

Table 4.6: The mean square error of reconstructed measurements with maximum likelihood (ML) and least squares (LS), respectively.

The MSE of the maximum likelihood (ML) estimate and the least squares (LS) estimate is illustrated in Table 4.6. As we can see that ML outperforms LS on reconstructing normal measurements.

4.7 Summary

In this chapter, a new design criterion known as optimal structured residuals for single faulty sensor identification is proposed. To maximize fault isolation ability, a set of optimal structured directions is designed. Each of them is insensitive to one of faults while being most sensitive to one of remaining ones. Fault isolability analysis shows that the column norms of the model matrix and the angles between two columns of the model matrix play a vital role in fault isolability. After identifying the faulty sensor, normal measurements are reconstructed through the maximum likelihood estimation. The application to an industrial boiler process demonstrates that the proposed OSR approach has superior fault identification ability.

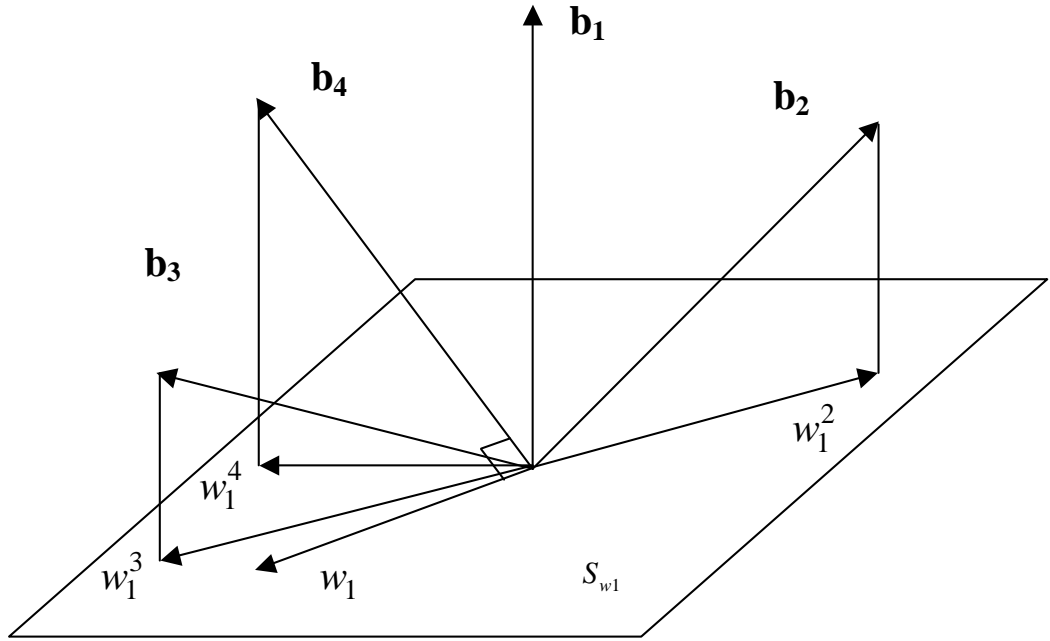


Figure 4.1: w_i^j ($j = 2, 3, 4$) is the projection of corresponding b_j ($j = 2, 3, 4$) to S_{w_1} , while w_1 may be orthogonal or nearly orthogonal to b_4 .

Chapter 5

An Optimal Structured Residuals Approach for Multidimensional Faulty Sensor Diagnosis

5.1 Introduction

Fault detection and isolation (FDI) in engineering systems are of great practical significance. The systems concerned encompass a broad spectrum of human made machinery, including industrial production facilities and household appliances. The early detection and diagnosis of faults are critical in avoiding product deterioration, performance degradation, major damage to the equipment and damage to human health or even loss of lives. The traditional approaches to fault detection and diagnosis involve the limit checking of some key variables or the application of redundant sensors (physical redundancy). Over the last two decades, fault detection and diagnosis have gained increasing consideration world-wide. This development was mainly stimulated by the trend of automation towards more complexity and the growing demand for higher security of control systems.

Advanced methods can be divided into two categories as qualitative model (knowledge model) based approach and quantitative model based approach. The objective of qualitative model based approach is to identify the

symptoms corresponding to the observations of the process that can be used for a fault decision on the basis of the knowledge redundancy, such as neural network, expert system and fuzzy logic. In the field of the quantitative model based approach, a strong impetus comes from the side of modern control theory that has brought forth mathematical modeling, state estimation and parameter identification that have been made feasible by the progress of modern computer technology.

The model based approach can be traced to chemical process control, where the traditional material and energy balance calculations evolved into systematic data reconciliation and the detection of gross errors. The work in this area is reviewed thoroughly in Crowe [13]. Another root can be traced to aerospace related research, this effort leads to the fundamental formulation of parity relation concepts [12]. An important related activity is due to Gertler and coworkers [20, 21], who try to diagnose faults by designing structured residuals that are insensitive to a particular subset of faults. In parallel to the above efforts, several researchers were looking into the possibility of applying Kalman filters [33, 41] and diagnostic observers [19, 49] to fault detection and isolation problem. In the area of fault detection and isolation by parameter estimation, substantial work has been done by Isermann and colleagues [28]. An important related activity is due to Basseville and coworkers, concerning the detection of small parametric faults by the statistical analysis of residuals obtained over extended sets of observations [4].

In this chapter, we extend the optimal structured residuals (OSR) ap-

proach in Chapter 4 to the multidimensional fault case. To maximize fault fault isolation ability, a matrix of optimal structured residuals have been designed. Each of them is insensitive to one subset of faults while being most sensitive to one of the remaining ones. The maximum of all structured residuals in each row is then selected for fault isolation. Faults occurred in dynamic systems can be considered as well using an extended state space model or dynamic principal component analysis.

The rest of this chapter is organized as follows. In Section 5.2, the fault detection and isolation problem for dynamic systems is briefly introduced. Multidimensional fault diagnosis with optimal structured residuals is presented in detail in Section 5.3. Necessary and sufficient conditions for fault isolation are investigated in Section 5.4. In Section 5.5, the utility of the proposed algorithm is illustrated using data from a 4×4 dynamic system. Section 5.6 conclude the chapter.

5.2 Fault diagnosis for dynamic systems

5.2.1 Problem formulation

In this chapter, we assume that the system can be written in innovations form as

$$x_{k+1} = Ax_k + Bu_k + Ke_k \quad (5.1a)$$

$$y_k = Cx_k + Du_k + e_k \quad (5.1b)$$

where $y_k \in R^{n_y}$, $x_k \in R^n$, $u_k \in R^{n_u}$, and $e_k \in R^{n_y}$ are the system output, state, input, and innovation, respectively. A , B , C and D are system matrices with appropriate dimensions. K is the Kalman filter gain. The system described by (5.1) can also be presented as

$$y_k = G(q)u_k + H(q)e_k \quad (5.2)$$

where $G(q) = C(qI - A)^{-1}B + D$, and $H(q) = C(qI - A)^{-1}K + I$.

Based on the state space description in (5.1), we can obtain

$$Y_f = \Gamma_f X_k + H_f U_f + G_f E_f \quad (5.3)$$

where the subscripts f denote future horizon. The extended observability matrix is

$$\Gamma_f = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{f-1} \end{bmatrix} \in \mathfrak{R}^{n_y f \times n} \quad (5.4)$$

and H_f and G_f are Toeplitz matrices:

$$H_f = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-2}B & CA^{f-3}B & \cdots & D \end{bmatrix} \in \mathfrak{R}^{n_y f \times n_u f}$$

$$G_f = \begin{bmatrix} I & 0 & \cdots & 0 \\ CK & I & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ CA^{f-2}K & CA^{f-3}K & \cdots & I \end{bmatrix} \in \mathfrak{R}^{n_y f \times n_y f}$$

The input and output data are arranged in the following Hankel form:

$$U_f = \begin{bmatrix} u_k & u_{k+1} & \cdots & u_{k+N-1} \\ u_{k+1} & u_{k+2} & \cdots & u_{k+N} \\ \vdots & \vdots & \ddots & \vdots \\ u_{k+f-1} & u_{k+f} & \cdots & u_{k+f+N-2} \end{bmatrix} \in \mathfrak{R}^{n_u f \times N}$$

$$Y_f = \begin{bmatrix} y_k & y_{k+1} & \cdots & y_{k+N-1} \\ y_{k+1} & y_{k+2} & \cdots & y_{k+N} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k+f-1} & y_{k+f} & \cdots & y_{k+f+N-2} \end{bmatrix} \in \mathfrak{R}^{n_y f \times N}$$

Defining

$$Z_f = \begin{bmatrix} Y_f \\ U_f \end{bmatrix} \in \mathfrak{R}^{(n_y+n_u)f \times N} \quad (5.5)$$

we can rewrite (5.3) as

$$\begin{bmatrix} I & -H_f \end{bmatrix} Z_f = \Gamma_f X_k + G_f E_f \quad (5.6)$$

We denote $\Gamma_f^\perp \in \mathfrak{R}^{n_y f \times (n_y f - n)}$ as the orthogonal complement of Γ_f such that

$$(\Gamma_f^\perp)^T \Gamma_f = 0 \quad (5.7)$$

Pre-multiplying (5.6) by $(\Gamma_f^\perp)^T$ leads to

$$(\Gamma_f^\perp)^T \begin{bmatrix} I & -H_f \end{bmatrix} Z_f = (\Gamma_f^\perp)^T G_f E_f \quad (5.8)$$

After defining

$$B_f \equiv (\Gamma_f^\perp)^T \begin{bmatrix} I & -H_f \end{bmatrix} \in \mathfrak{R}^{(n_y f - n) \times (n_y + n_u) f} \quad (5.9)$$

$$V_f \equiv (\Gamma_f^\perp)^T G_f E_f \in \mathfrak{R}^{(n_y f - n) \times N} \quad (5.10)$$

we obtain

$$B_f Z_f = V_f \quad (5.11)$$

Equation (5.11) can also be presented as

$$B_f z_f(t) = v_f(t) \quad (5.12)$$

where $z_f(t)$ and $v_f(t)$ are column vectors of Z_f and V_f respectively.

If the faulty input/output data are represented as

$$z_f(t) = z_f^*(t) + \Xi_i f_i(t) \quad (5.13)$$

we can obtain

$$v_f(t) = B_f z_f^*(t) + B_f \Xi_i f_i(t) = v_f^*(t) + \Phi_i f_i(t) \quad (5.14)$$

where $f_i(t) \in \mathbb{R}^{l_s \times 1}$ is the vector of the fault magnitude, l_s is the dimension of the fault and $\Xi_i \in \mathbb{R}^{(n_y+n_u)f \times l_s}$ is the matrix of fault directions with appropriate dimensions. $\Phi_i = B_f \Xi_i$ is the fault direction matrix on the model residual $v_f(t)$. Vectors $z_f^*(t)$ and $v_f^*(t)$ stand for measurements and model residuals under normal condition, respectively. The objective of fault detection and isolation is to detect the onset of the fault, identify fault direction matrix, Ξ_i , and estimate the fault magnitude, $f_i(t)$, from faulty measurements. One shall notice that with this formulation even a single sensor fault is multi-dimensional.

5.2.2 Temporal redundancy and its relationship with dynamic PCA

Equations (5.11) and (5.12) can be regarded as the parity relationships based on temporal redundancy. On the other hand, principal component analysis (PCA) has been widely used in process monitoring applications [53]. The steady-state PCA approach can be extended to dynamic principal component analysis (DPCA) [36] for data from dynamic systems. The relationship between PCA and static redundancy has been discussed in [55]. In this section, we discuss the relationship between temporal redundancy and DPCA following [38].

The DPCA can be regarded as performing PCA of the following Hankel matrix

$$Z_l^T = \begin{bmatrix} z_k^T & z_{k+1}^T & \cdots & z_{k+l-1}^T \\ z_{k+1}^T & z_{k+2}^T & \cdots & z_{k+l}^T \\ \vdots & \vdots & \ddots & \vdots \\ z_{k+N-1}^T & z_{k+N}^T & \cdots & z_{k+l+N-2}^T \end{bmatrix} \quad (5.15)$$

where $Z_l^T \in \Re^{N \times (n_y + n_u)l}$, z_k is the process measurement and l is the number of lags in DPCA.

After PCA decomposition of Z_l^T , we obtain

$$Z_l^T = \hat{T} \hat{P}^T + \tilde{T} \tilde{P}^T \quad (5.16)$$

where \hat{T} is the score matrix and \hat{P} is the loading matrix. The decomposition is made such that $\begin{bmatrix} \hat{T} & \tilde{T} \end{bmatrix}$ is orthogonal and $\begin{bmatrix} \hat{P} & \tilde{P} \end{bmatrix}$ is orthonormal. Multiplying \tilde{P} on the right hand side of (5.16), we obtain

$$Z_l^T \tilde{P} = \tilde{T} \quad (5.17)$$

Transposing the above matrix, we obtain

$$\tilde{P}^T Z_l = \tilde{T}^T \quad (5.18)$$

Comparing (5.18) with (5.11), we have following observations:

1. \tilde{P}^T is similar to $(\Gamma_f^\perp)^T \begin{bmatrix} I & -H_f \end{bmatrix}$.
2. l in DPCA is equivalent to f in extended state space model.

From the above analysis, we can see that the temporary redundancy can be established not only from the extended state space model but also from DPCA as well. However, to achieve consistent estimate of B_f from \tilde{P} , the direct DPCA approach in (5.16) has to be modified. Li and Qin [38] have proposed an indirect DPCA approach that uses instrumental variables to achieve consistency.

5.3 Multidimensional faults diagnosis with optimal structured residuals

5.3.1 Optimal structured residuals approach

The structured residual approach is introduced by Gerlter and Singer [21]. In their work, the structured residual, $r_i(t)$, is defined as

$$r_i(t) = w_i^T v_f^*(t) + w_i^T \Phi_i f_i(t) \quad (5.19)$$

subject to

$$w_i^T \Phi_i = 0$$

where w_i is the structured residual direction. Φ_i is assumed to have unit column norm without loss of generality. With this design criterion, the structured residual is insensitive to the i^{th} fault. As pointed out by Qin and Li [55], the structured residual approach does not maximize the fault isolation ability. Based on this analysis, they propose structured residual approach with maximized sensitivity (SRAMS).

In our recent work [39], we analyze the SRAMS approach in detail and point out the possibility of misidentifying faults. We also propose a new optimal structured residuals (OSR) criterion for fault diagnosis. For the multidimensional fault, the optimal structured directions can be extended as follows:

$$\begin{aligned} w_i^j &= \max_{\|w_i^j\|=1} \|\Phi_j^T w_i^j\|^2 \\ i &= 1, 2, \dots, n_f \\ j &= 1, 2, \dots, (i-1), (i+1), \dots, n_f \end{aligned} \tag{5.20}$$

subject to

$$w_i^{jT} \Phi_i = 0 \tag{5.21}$$

where n_f is the number of faults under consideration.

To derive an explicit expression of w_i^j , we choose w_i^j as

$$w_i^j = \Pi_{\Phi_i}^\perp z_i^j \tag{5.22}$$

such that (5.21) is satisfied, where z_i^j is a vector with appropriate dimension. $\Pi_{\Phi_i}^\perp$ is the projection matrix on the orthogonal complement of Φ_i .

After introducing a Lagrange multiplier, the objective function can be defined as

$$J = \|\Phi_j^T \Pi_{\Phi_i}^\perp z_i^j\|^2 + \lambda(1 - \|\Pi_{\Phi_i}^\perp z_i^j\|^2)$$

Differentiating the objective function with respect to z_i^j and with the help of (5.22), we can obtain

$$\Pi_{\Phi_i}^\perp \Phi_j \Phi_j^T \Pi_{\Phi_i}^\perp z_i^j = \lambda \Pi_{\Phi_i}^\perp z_i^j$$

or

$$\Pi_{\Phi_i}^\perp \Phi_j \Phi_j^T \Pi_{\Phi_i}^\perp w_i^j = \lambda w_i^j$$

Therefore, w_i^j is the eigenvector of $\Pi_{\Phi_i}^\perp \Phi_j \Phi_j^T \Pi_{\Phi_i}^\perp$ corresponding to the largest eigenvalue. Equivalently, w_i^j is the left singular vector of $\Pi_{\Phi_i}^\perp \Phi_j$ corresponding to the largest singular value.

After obtaining the structured direction, the structured residual for faulty data can be defined as:

$$r_i^j(t) = w_i^{jT} v_f(t) = w_i^{jT} [v_f^*(t) + \Phi_g f_g(t)] \quad (5.23)$$

From the definition of the structured residual, we can see that

$$r_i^j(t) = \begin{cases} w_i^{jT} v_f^*(t) & i = g, \\ w_i^{jT} v_f^*(t) + w_i^{jT} \Phi_g f_g(t) & i \neq g. \end{cases} \quad (5.24)$$

From (5.24), we can obtain a structured residuals matrix $R(t) \in \mathbb{R}^{n_f \times n_f}$.

$$R(t) = \begin{bmatrix} - & r_1^2(t) & r_1^3(t) & \cdots & r_1^{n_f}(t) \\ r_2^1(t) & - & r_2^3(t) & \cdots & r_2^{n_f}(t) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{n_f-1}^1(t) & \cdots & r_{n_f-1}^{n_f-2}(t) & - & r_{n_f-1}^{n_f}(t) \\ r_{n_f}^1(t) & r_{n_f}^2(t) & \cdots & r_{n_f}^{n_f-1}(t) & - \end{bmatrix} \quad (5.25)$$

We choose the one with maximum absolute value among each row of $R(t)$ for fault diagnosis:

$$r_i(t) = \max_j |r_i^j(t)| \quad i = 1, \dots, n_f \quad (5.26)$$

If fault g occurs, Equation (5.26) picks the one with maximum absolute value among each row of $R(t)$ as

$$r_i(t) = r_i^g(t) = w_i^{gT} [v_f^*(t) + \Phi_g f_g(t)] \quad (5.27)$$

5.3.2 Fault identification indices

After obtaining structured residuals as defined in (5.26), one needs to design appropriate statistical inference for fault diagnosis. In this subsection, we discuss the deterministic fault and the stochastic fault separately along with identification indices defined by Qin and Li [55].

For deterministic faults, such as bias or drifting, the distribution of $r_i^j(t)$ in (5.24) is

$$r_i^j(t) \sim \begin{cases} N(0, w_i^{jT} R_e w_i^j) & i = g, \\ N(w_i^{jT} \Phi_g \bar{f}_g, w_i^{jT} R_e w_i^j) & i \neq g. \end{cases} \quad (5.28)$$

where \bar{f}_k is the bias of the fault, and R_e is the covariance matrix of normal residuals $v_f^*(t)$. After picking up $r_i(t)$ with (5.26), under normal measurements we have

$$\frac{r_i^2(t)}{w_i^{kT} R_e w_i^k} \sim \chi^2(1),$$

or

$$\frac{r_i^2(t)}{w_i^{kT} R_e w_i^k} \leq \chi_\alpha^2(1)$$

with confidence level α . Therefore, one can define a structured fault identification index as

$$I_{SR}^i(t) = \frac{r_i^2(t)}{w_i^{kT} R_e w_i^k \chi_\alpha^2(1)}. \quad (5.29)$$

Under normal conditions $I_{SR}^i < 1$; if there is a fault in the g^{th} sensor $I_{SR}^i > 1$ except for $i = g$. One can also apply an EWMA filter to structured residuals to obtain a filtered structured residual index (I_{FSR}^i) [55].

For stochastic faults, such as precision degradation, the distribution of $r_i^j(t)$ in (5.24) is

$$r_i^j(t) \sim \begin{cases} N(0, w_i^{jT} R_e w_i^j) & i = g, \\ N(0, w_i^{jT} R_d w_i^j) & i \neq g. \end{cases}$$

where R_d is the covariance matrix of degraded measurements. Assuming that the stochastic fault is independent of $v_f^*(t)$, the covariance R_d can be derived from (5.14) as follows

$$R_d = R_e + \Phi_i R_v \Phi_i^T \quad (5.30)$$

where R_v is the covariance of the fault. As proposed by Qin and Li [55], one can apply cumulative variance (V_{sum}) index to identify variance change.

$$V_{sum}^i(t) = \sum_{l=t-T}^t (r_i(l) - \hat{\mu}_i)^2$$

such that

$$\hat{\mu}_i = \frac{\sum_{l=t-T}^t r_i(l)}{T+1}$$

where T is the moving window size. After picking up $r_i(t)$ with (5.26), with no fault we have,

$$\frac{V_{sum}^i(t)}{w_i^{kT} R_e w_i^k} \sim \chi^2(T)$$

Therefore one can define a V_{sum} index as

$$I_{V_{sum}}^i(t) = \frac{V_{sum}^i(t)}{w_i^{kT} R_e w_i^k \chi_\alpha^2(T)} \quad (5.31)$$

If $I_{V_{sum}}^i(t) > 1$ for all i but $i = g$, the g^{th} sensor has a variance fault with confidence level α .

After detecting the occurrence of faults, a faulty sensor can be identified as follows.

1. Based on the process model matrix B_f , design optimal structured directions w_i^j off-line.
2. From the process residual $v_f(t)$, calculate optimal structured residuals $r_i(t)$ and corresponding identification indices.
3. If the i^{th} index is less one while all other indices are greater than one, the i^{th} sensor is faulty.

5.4 Fault isolability

The fault isolability issue for quasi-steady-state process has been discussed by Lin and Qin [39]. Here we extend our analysis to the multidimen-

sional case. As we can see, for multidimensional faults, if the range space of Φ_i and the range space of Φ_j are totally overlapped, the i^{th} fault and the j^{th} fault are not isolable. Therefore the necessary condition for the i^{th} fault to be isolable from the j^{th} is that the range space of Φ_i and the range space of Φ_j are not totally overlapped. However when the range spaces of i^{th} fault and j^{th} fault are very closed to each other, say principal angles [65] between two of them are close to zero, they are difficult to isolate, unless the fault magnitude is sufficient large.

We can discuss the sufficient condition for the multidimensional fault isolability, which is how large the fault magnitude should be to guarantee fault isolability. In this section, we investigate the sufficient condition for multidimensional deterministic faults isolability. When the g^{th} fault occurs, from (5.27) and (5.4) we have

$$r_i(t) = r_i^g(t) \sim \begin{cases} N(0, w_i^{gT} R_e w_i^g) & i = g, \\ N(w_i^{gT} \Phi_g \bar{f}_g, w_i^{gT} R_e w_i^g) & i \neq g. \end{cases}$$

Therefore, we can obtain [32],

$$Y_i(t) = \frac{r_i^2(t)}{w_i^{gT} R_e w_i^g} \sim \begin{cases} \chi^2(1) & i = g, \\ \chi^2(1, \lambda_i) & i \neq g. \end{cases} \quad (5.32)$$

where $\chi^2(1, \lambda_i)$ is the non-central χ^2 with one degree of freedom and non-centrality

$$\lambda_i = \frac{(w_i^{gT} \Phi_g \bar{f}_g)^2}{w_i^{gT} R_e w_i^g}$$

To guarantee that the i^{th} fault is isolable from the g^{th} fault with some

confidence level β , we should have

$$P\{I_{SR}^i(t) = \frac{Y_i(t)}{\chi_\alpha^2(1)} > 1\} = 1 - P[Y_i(t) \leq \chi_\alpha^2(1)] = \beta$$

Therefore, if we find a critical value, λ_{ic} , such that

$$\chi_{1-\beta}^2(1, \lambda_{ic}) = \chi_\alpha^2(1),$$

then $I_{SR}^i(t) > 1$ with confidence level β for $\lambda_i > \lambda_{ic}$. Therefore, from (5.32) we can derive the sufficient condition for deterministic fault isolability as

$$\|\bar{f}_g\| \geq \frac{(w_i^{gT} Rew_i^g \lambda_{ic})^{\frac{1}{2}}}{\|\Phi_g^T\| \cos \gamma} \quad (5.33)$$

where γ is the angle between \bar{f}_g and $w_i^{gT} \Phi_g$. From the above equation, we can see that two important factors are $\|\Phi_g^T\|$ and γ .

5.5 Simulation example: a 4×4 dynamic process

To test the effectiveness of the proposed method we simulate a second order 4×4 dynamic process. The system matrices are:

$$x_{k+1} = \begin{bmatrix} 0.67 & 0.67 \\ -0.67 & 0.67 \end{bmatrix} x_k \quad (5.34a)$$

$$\begin{aligned} & + \begin{bmatrix} -0.4326 & 0.1253 & -1.1465 & 1.1892 \\ -1.6656 & 0.2877 & 1.1909 & -0.0376 \end{bmatrix} u_k \\ & + \begin{bmatrix} 0.0654 & 0.0836 & -0.0465 & 0.1727 \\ -0.1147 & 0.0593 & 0.0426 & -0.1858 \end{bmatrix} e_k \\ y_k & = \begin{bmatrix} 0.3273 & -0.5883 \\ 0.1746 & 2.1832 \\ -0.1867 & -0.1364 \\ 0.7258 & 0.1139 \end{bmatrix} x_k \\ & + \begin{bmatrix} 1.0668 & 0.2944 & -0.6918 & -1.4410 \\ -0.0593 & -1.3362 & 0.8580 & 0.5711 \\ -0.0956 & 0.7143 & 1.2540 & -0.3999 \\ -0.8323 & 1.6236 & -1.5937 & 0.6900 \end{bmatrix} u_k \\ & + e_k \quad (5.34b) \end{aligned}$$

The process inputs are Gaussian white noise sequence with identity covariance matrix, the measure noises are Gaussian white noise sequence as well with identity covariance matrix. 500 samples of the input and output data are collected. The basic statistics of the data are listed in Table 5.1. The data are scaled to zero mean and unit variance, and the process model B_f is built using DPCA with $l = 4$. The number of principal components is determined by the best reconstruction criterion [54]. The process model B_f is a 14×32 matrix.

In this work, two kinds of faults are simulated: bias and precision

Sensor No.	Minimum	Maximum	Mean	Stand Deviation
1	-11.07	11.91	0.2467	3.7788
2	-35.51	32.20	-0.4773	10.798
3	-5.839	6.008	0.1201	2.0616
4	-13.36	11.58	-0.0247	4.2254

Table 5.1: Basic statistics for the outputs of the 4×4 process

degradation, which can be represented as

$$f_i(t) = C_1 \quad (5.35)$$

$$f_i(t) \sim N(0, C_2) \quad (5.36)$$

where C_1 and C_2 are constants for different faults, and f_i is the fault magnitude. Bias and precision degradation faults represent changes in mean and variance, which are suitable for EWMA filtered structured residual index and cumulative variance index respectively. To demonstrate the fault identification ability, single sensor faults with different fault magnitudes are introduced as the 200th sample. Since the purpose of this work is not to demonstrate the fault detection ability, we assume that the fault is detected at the 250th sample.

To identify faults, the filtered structured residuals index (I_{FSR}) with EWMA filter coefficient 0.98 and the cumulative variance index ($I_{V_{sum}}$) with a moving window size of 20 are used in this work. They are applied to identify bias and precision degradation, respectively.

With the process model $B_f \in \Re^{14 \times 32}$, the structured residuals matrix

$R(t)$ is a 8×8 matrix. Every element of $R(t)$ is calculated with (5.23).

We then pick up $r_i(t)$ according to (5.26) and calculate the corresponding fault identification indices. The single sensor for the dynamic system can be identified as discuss at the end of Subsection 5.3.2.

The fault identification results for a single biased sensor fault with OSR are shown in Table 5.2. Fault magnitudes are calculated as percentage of the standard deviation of the corresponding sensor measurements. Normal sensor measurements and faulty measurements with a bias fault at the 4th sensor with fault magnitude as 40% of the standard deviation of the corresponding measurements are shown in Fig. 5.1, the average I_{FSR} and I_{Vsum} for the same case are shown in Fig. 5.2. From the results we can see the OSR approach can correctly identify the biased sensor at small fault magnitudes, which cannot be identified by direct visualization.

	Faulty Sensor			
	1	2	3	4
0.65%	—	—	—	—
3.0%	✓	—	—	—
7.0%	✓	✓	✓	—
40.0%	✓	✓	✓	✓

Table 5.2: Fault identification results for the single biased sensor with I_{FSR} through OSR approach. ✓ means the faulty sensor can be correctly identified, while — means the faulty sensor cannot be identified.

The fault identification results for a single precision degradation sensor with OSR are shown in Table 5.3. Fault magnitudes are calculated as

the percentage of the standard deviation of the corresponding sensor measurements. Normal sensor measurements and faulty measurements with a precision degradation at the 1st sensor with the fault magnitude as 25.0% of the standard deviation of the corresponding measurements are shown in Fig. 5.3, the average I_{FSR} and I_{Vsum} for the same case are shown in Fig. 5.4. From the results we can see the OSR approach can correctly identify the faulty sensor at small fault magnitude, which can not be identified by direct visualization.

	Faulty Sensor			
	1	2	3	4
9.0%	—	—	—	—
17.0%	—	✓	—	—
25.0%	✓	✓	✓	✓

Table 5.3: Fault identification results for the single precision degraded sensor with V_{sum} through OSR approach. ✓ means the faulty sensor can be correctly identified, while — means the faulty sensor cannot be identified.

5.6 Summary

In this chapter, an optimal structured residuals (OSR) approach for multidimensional fault diagnosis is proposed. Faults in dynamic processes can be handled with the extended state space model or dynamic PCA. The relationship between temporary redundancy and dynamic PCA allows us to use a dynamic PCA model for fault detection and diagnosis. A multi-input-multi-output simulation study shows that the proposed approach can identify faulty sensors with small fault magnitudes.

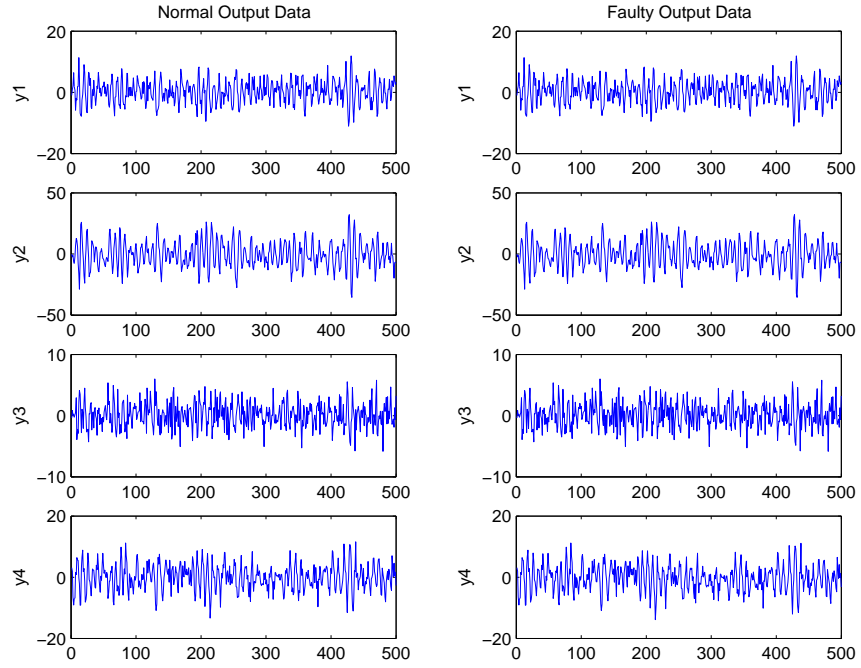


Figure 5.1: The output measurement for bias fault at the 4th sensor with fault magnitude as 40% of the standard deviation of the corresponding measurements

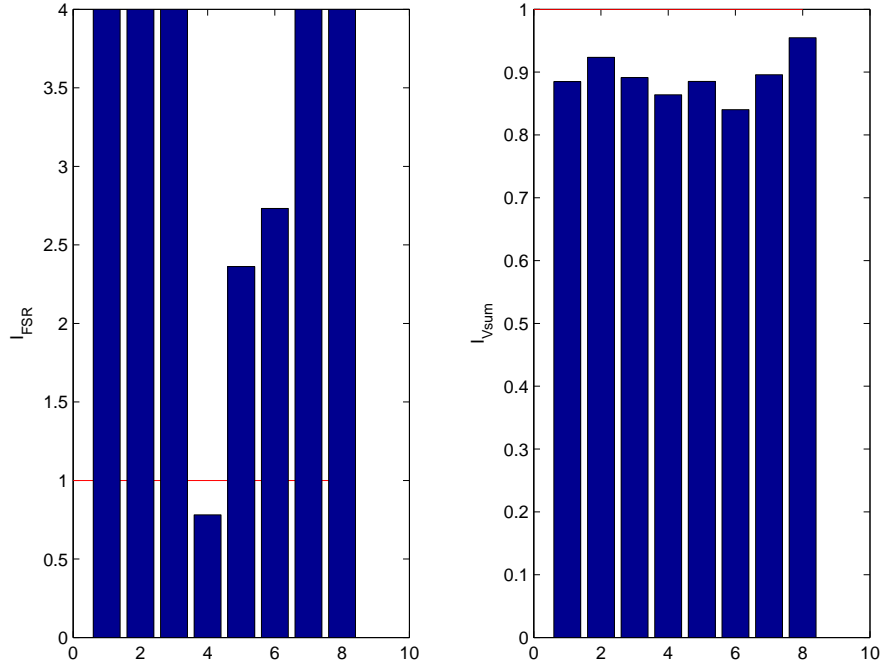


Figure 5.2: The average I_{FSR} and I_{Vsum} for bias fault at the 4th sensor with fault magnitude as 40% of the standard deviation of the corresponding measurements

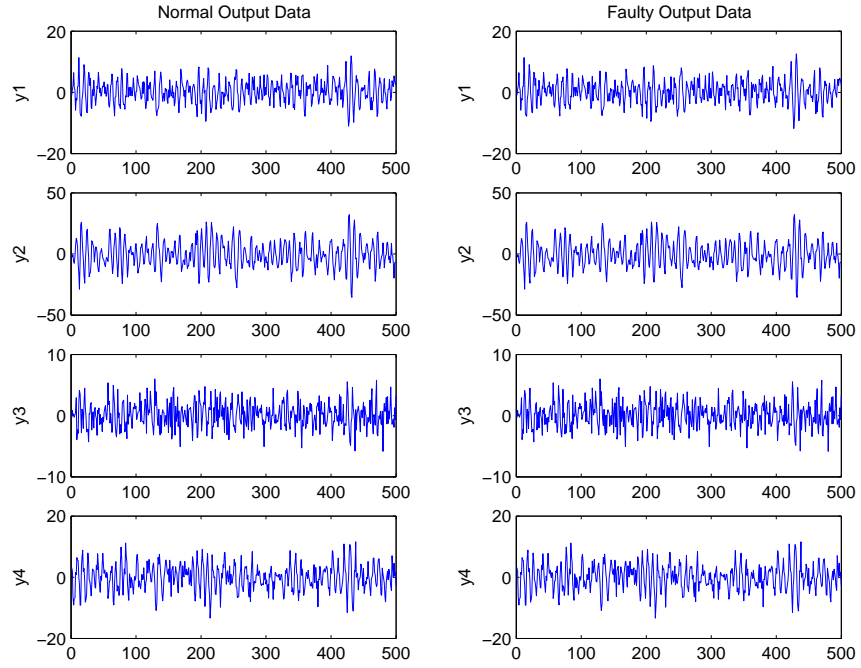


Figure 5.3: The output measurement for precision degradation fault at the 1^{th} sensor with fault magnitude as 25.0% of the standard deviation of the corresponding measurements

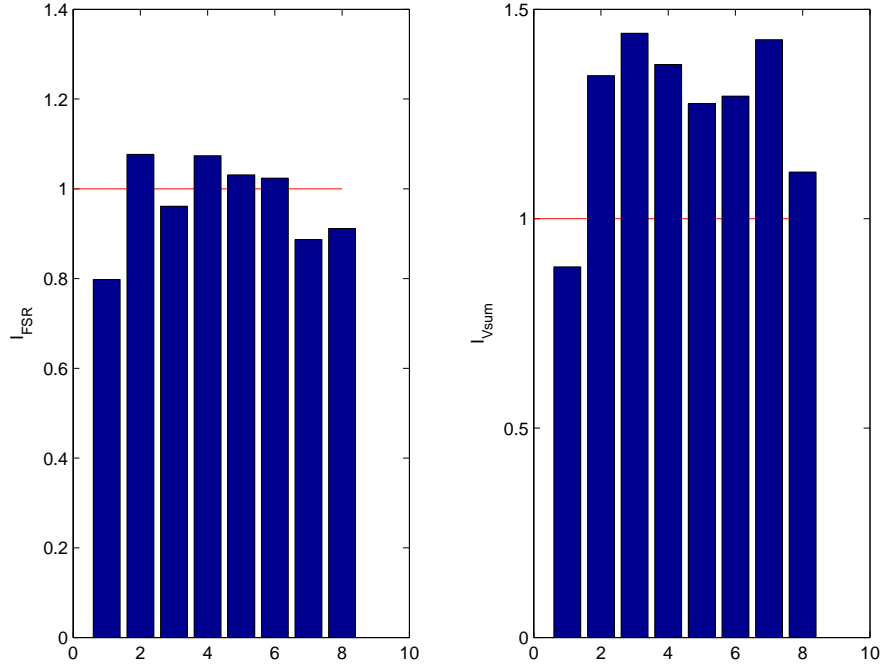


Figure 5.4: The average I_{FSR} and I_{Vsum} for precision degradation fault at the 1th sensor with fault magnitude as 25.0% percentage of the standard deviation of the corresponding measurements

Chapter 6

Conclusions and Future Research Suggestions

The main purpose of this work is to investigate disadvantages of traditional subspace identification algorithms and structured residuals approaches for process modeling and diagnosis. The contributions of present work are listed in Section 6.1, while suggestions for future research effort are presented in Section 6.2.

6.1 Contributions

The contributions of this research are listed below.

1. Novel subspace identification methods (SIMs) with enforced causal models are implemented. It has been shown that proposed algorithm has lower estimation variance comparing to traditional SIMs. Meanwhile the consistency analysis shows that the proposed algorithms are consistent under certain assumptions.
2. The feasibility of closed-loop subspace identification is investigated. Novel closed-loop subspace identification methods with innovation estimation are proposed. The new algorithms are shown to be consistent under

closed-loop condition, while the traditional SIMs fail to provide consistent estimates.

3. In this work, another closed-loop SIM referred to as the "whitening filter approach" is analyzed. The similarity and difference between the "whitening filter approach" and the "innovation estimation approach" are investigated. It turns out that although they are based on different representations of state space model. Both of them can be implemented through multi-stage least squares.
4. A new optimal structured residuals (OSR) approach for unidirectional faults diagnosis is proposed. To maximize fault identification ability, a set of optimal structured residuals have been designed. Each of them is insensitive to one of the faults while being most sensitive to one of the remaining ones. The fault isolability analysis shows that the column norms of the model matrix and the angles between two columns play a vital role in fault isolability.
5. The OSR for unidirectional fault diagnosis is extended to multidimensional fault diagnosis. A set of optimal structured residuals have been designed. Each of them is insensitive to one subset of faults while being most sensitive to one of the remaining ones. The fault isolability analysis shows that the 2-norm of the fault direction matrix and the angle between fault magnitude vector and certain linear transformation of structured residuals are important for fault isolability.

6.2 Future research suggestions

In this section, we point out some future research directions for subspace identification algorithms as well as the optimal structured residuals approach with applications in process modeling and diagnosis areas.

1. Closed-loop subspace identification algorithms are very attractive from a practical point of view. In this work we have proposed closed-loop SIMs with innovation estimation. As there are other closed-loop identification algorithms available, the advantage and disadvantage of them are worth further investigation from both theoretical and engineering perspectives.
2. The work in this dissertation focuses on subspace identification for linear time invariant systems, while in practice a lot of systems have either time variant or nonlinear characteristics. How to characterize such properties from data and identify a model for systems with such characteristics are very interesting topics.
3. The optimal structured residuals approach proposed in this work only deals with the sensor fault. How to apply the approach to more complicated situations such as closed-loop faults diagnosis is worth further investigation.

Appendices

Appendix A

Proof of Theorem 3.2.2

From (3.17) we can conclude that,

$$\begin{bmatrix} Z_p \\ U_{i-1} \\ E_{i-1} \end{bmatrix}^\dagger = \begin{bmatrix} Z_p \\ U_{i-1} \\ E_{i-1} \end{bmatrix}^T \left[\begin{bmatrix} Z_p \\ U_{i-1} \\ E_{i-1} \end{bmatrix} \begin{bmatrix} Z_p \\ U_{i-1} \\ E_{i-1} \end{bmatrix}^T \right]^{-1}$$

in (3.18). Therefore, the estimate of $\Gamma_{fi}L_z$, H_{fi}^- and G_{fi}^- is consistent if and only if

1. E_{fi} is uncorrelated with Z_p , U_{i-1} and E_{i-1} in (3.16).
2. $\begin{bmatrix} Z_p \\ U_{i-1} \\ E_{i-1} \end{bmatrix}$ has full row rank (f.r.r.).

The first condition is satisfied as mentioned in Subsection 3.2.2. Here we provide a proof for the second condition by induction.

For $i = 2$, we can obtain E_{f1} from (3.16) as

$$E_{f1} = Y_{f1} - \Gamma_{f1}L_zZ_p$$

Therefore,

$$\begin{bmatrix} Z_p \\ U_1 \\ E_1 \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ M_1 & 0 & I \end{bmatrix} \begin{bmatrix} Z_p \\ U_1 \\ Y_1 \end{bmatrix}$$

where $M_1 = -\Gamma_{f1}L_z$. From the results of Lemma 1, we know that χ_k is persistently exciting of any order. Therefore, we can conclude that $\begin{bmatrix} Z_p^T & U_1^T & E_1^T \end{bmatrix}^T$ has f.r.r. Now we assume that, for $i = k + 1$,

$$\begin{bmatrix} Z_p \\ U_k \\ E_k \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ M_k & N_k & I \end{bmatrix} \begin{bmatrix} Z_p \\ U_k \\ Y_k \end{bmatrix}$$

where M_k and N_k are matrices with appropriate dimensions. Then for $i = k+2$, from (3.17), we obtain

$$\begin{aligned} E_{f(k+1)} &= Y_{f(k+1)} - \Gamma_{f(k+1)}L_z Z_p - H_{f(k+1)}^- U_k - G_{f(k+1)}^- E_k \\ &= Y_{f(k+1)} - (\Gamma_{f(k+1)}L_z + G_{f(k+1)}^- M_k) Z_p \\ &\quad - (H_{f(k+1)}^- + G_{f(k+1)}^- N_k) U_k - G_{f(k+1)}^- Y_k \end{aligned}$$

Therefore,

$$\begin{bmatrix} Z_p \\ U_k \\ U_{f(k+1)} \\ E_k \\ E_{f(k+1)} \end{bmatrix} = \begin{bmatrix} I & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \\ M_k & N_k & 0 & I & 0 \\ M_{k+1} & N_{1(k+1)} & 0 & N_{2(k+1)} & I \end{bmatrix} \begin{bmatrix} Z_p \\ U_k \\ U_{f(k+1)} \\ Y_k \\ Y_{f(k+1)} \end{bmatrix}$$

where

$$\begin{aligned} M_{k+1} &= -(\Gamma_{f(k+1)}L_z + G_{f(k+1)}^- M_k) \\ N_{1(k+1)} &= -(H_{f(k+1)}^- + G_{f(k+1)}^- N_k) \\ N_{2(k+1)} &= -G_{f(k+1)}^- \end{aligned}$$

Then from Lemma 3.2.1, we conclude that the RHS of the above equation is f.r.r..

Appendix B

Proof of Theorem 4.3.1

For the single fault case, w_i designed by SRAMS that is insensitive to the i^{th} fault is:

$$B_i^\circ B_i^{\circ T} w_i = \lambda w_i$$

where $B_i^\circ = (I - b_i^o b_i^{oT}) B^\circ$, and w_i is the eigenvector of $B_i^\circ B_i^{\circ T}$ corresponding to the largest eigenvalue.

Let $B' = [B\Xi_1, \dots, B\Xi_{j-1}, B\Xi_{j+1}, \dots, B\Xi_n]$, w'_i designed by SRAMS that is also insensitive to the i^{th} fault is

$$\tilde{B}_i^\circ \tilde{B}_i^{\circ T} w'_i = \lambda w'_i$$

where $\tilde{B}_i^\circ = (I - b_i^o b_i^{oT}) B'^\circ$ and w'_i is the eigenvector of $\tilde{B}_i^\circ \tilde{B}_i^{\circ T}$ corresponding to the largest eigenvalue. It is obvious that

$$B_i^\circ B_i^{\circ T} w_i = \tilde{B}_i^\circ \tilde{B}_i^{\circ T} w_i + (I - b_i^o b_i^{oT}) B^\circ \Xi_j \Xi_j^T B^{\circ T} (I - b_i^o b_i^{oT}) w_i$$

Noticing that w_i is orthogonal to b_i^o , therefore

$$(I - b_i^o b_i^{oT}) w_i = w_i$$

So $w'_i = w_i$ if and if only

$$(I - b_i^o b_i^{oT}) B^\circ \Xi_j \Xi_j^T B^{\circ T} w_i = 0 \quad \text{for } \forall w_i$$

which is equivalent to $\Xi_j^T B^T w_i = 0$, or w_i is orthogonal to the j^{th} fault as well. Therefore fault $B\Xi_i$ and $B\Xi_j$ are not isolable with each other.

Bibliography

- [1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, Inc., second edition, 1984.
- [2] W. F. Arnold and A. J. Laub. Generalized eigenproblem algorithms and software for algebraic Riccati equations. *Proc. IEEE*, 72:1746–1754, 1984.
- [3] K. J. Astrom and T. Bohlin. Numerical identification of linear dynamic systems from normal operating records. In *Proc. IFAC Symposium on Self-Adaptive Systems*, 1965.
- [4] M. Basseville. Information criteria for residual generation and fault detection and isolation. *Automatica*, 33:783–803, 1997.
- [5] M. Basseville and I. V. Nikiforov. *Detection of Abrupt Change - Theory and Applications*. Prentice-Hall, Englewood Cliffs, 1993.
- [6] D. Bauer and M. Jansson. Analysis of the asymptotic properties of the MOESP type of subspace algorithm. *Automatica*, 36:497–509, 2000.
- [7] D. Bauer and L. Ljung. Some facts about the choice of the weighting matrices in Larimore type of subspace algorithms. *Automatica*, 38:763–773, 2002.

- [8] A. Chiuso and G. Picci. The asymptotic variance of subspace estimates. *J. of Econometrics*, 118:257–291, 2004.
- [9] A. Chiuso and G. Picci. Consistency analysis of some closed-loop subspace identification methods. *Automatica*, 41:377–391, 2005.
- [10] Alessandro Chiuso and Giorgio Picci. Constructing the state of random processes with feedback. In *Proceedings of the 13th IFAC SYSID Symposium*, pages 881–886, Rotterdam, NL, Aug 2003.
- [11] C. T. Chou and M. Verhaegen. Subspace algorithms for the identification of multivariable dynamic errors-in-variables models. *Automatica*, 33:1857–1869, 1997.
- [12] E. Y. Chow and A. S. Willsky. Analytical redundancy and the design of robust failure detection systems. *IEEE Trans. Auto. Cont.*, 29:603–614, 1984.
- [13] C. Crowe. Data reconciliation-progress and challenge. *J. Process Control*, 6:89–98, 1996.
- [14] D. Di Ruscio. Combined deterministic and stochastic system identification and realization: DSR-a subspace approach based on observations. *Modeling, Identification and Control*, 17:193–230, 1996.
- [15] R. Dunia and S. J. Qin. Subspace approach to multidimensional fault identification and reconstruction. *AIChE J.*, 44:1813–1830, 1998.

- [16] R. Dunia, S. J. Qin, T. F. Edgar, and T. J. McAvoy. Identification of faulty sensors using principal component analysis. *AIChE J.*, 42:2797–2812, 1996.
- [17] U. Forssell and L. Ljung. Closed-loop identification revisited. *Automatica*, 35:1215–1241, 1999.
- [18] P. M. Frank. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy - a survey and some new results. *Automatica*, 3:459–474, 1990.
- [19] P. M. Frank and X. Ding. Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *J. Process Control*, 6:403–424, 1997.
- [20] J. Gertler. *Fault detection and diagnosis in engineering systems*. Marcel Dekker, Inc., 1998.
- [21] J. Gertler and D. Singer. A new structural framework for parity equation based failure detection and isolation. *Automatica*, 26:381–388, 1990.
- [22] G.H. Golub and C.F. Van Loan. *Matrix Computations*. Johns Hopkins, third edition, 1996.
- [23] T. Gustafsson. Subspace-based system identification: weighting and pre-filtering of instruments. *Automatica*, 38:433–443, 2002.

- [24] T. Gustafsson and B. D. Rao. Statistical analysis of subspace-based estimation of reduced-rank linear regression. *IEEE Trans. Signal Processing*, 50:151–159, 2002.
- [25] H. Hjalmarsson. From experiments to closed loop control. In *Proceedings of the 13th IFAC SYSID Symposium*, pages 1–14, Rotterdam, NL, Aug 2003.
- [26] B. L. Ho and R. E. Kalman. Effective construction of linear state-variable models from input-output functions. *Regelungstechnik*, 12:545–548, 1965.
- [27] B. Huang, S. X. Ding, and S. J. Qin. Closed-loop subspace identification: an orthogonal projection approach. *J. Process Control*, 15:53–66, 2005.
- [28] R. Isermann. Fault diagnosis of machines via parameter estimation and knowledge processing. *Automatica*, 29:815–835, 1993.
- [29] M. Jansson. Subspace identification and ARX modelling. In *Proceedings of the 13th IFAC SYSID Symposium*, Rotterdam, NL, Aug 2003.
- [30] M. Jansson and Bo Wahlberg. A linear regression approach to state-space subspace system identification. *Signal Processing*, 52:103–129, 1996.
- [31] M. Jansson and Bo Wahlberg. On consistency of subspace methods for system identification. *Automatica*, 34:1507–1519, 1998.
- [32] N. L. Johnson and S. Kotz. *Distributions in Statistics: Continuous Univariate Distributions-2*. John Wiley and Sons, Inc., 1970.

- [33] J. Y. Keller. Fault isolation filter design for linear stochastic system. *Automatica*, 35:1701–1706, 1999.
- [34] T. Knudsen. Consistency analysis of subspace identification methods based on linear regression approach. *Automatica*, 37:81–89, 2001.
- [35] M. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.*, 37:233–243, 1991.
- [36] W. Ku, R. Storer, and C. Georgakis. Disturbance detection and isolation by dynamic principal component analysis. 30:179–196, 1995.
- [37] W. E. Larimore. Canonical variate analysis in identification, filtering and adaptive control. In *IEEE Conference on Decision and Control*, pages 596–604, Dec 1990.
- [38] W. Li and S. J. Qin. Consistent dynamic PCA based on errors-in-variables subspace identification. *J. Process Control*, 11:661–678, 2001.
- [39] W. Lin and S. J. Qin. An optimal structured residual approach for improved faulty sensor diagnosis. *Ind. Eng. Chem. Res. In Press*, 2005.
- [40] W. Lin, S. J. Qin, and Lennart Ljung. On consistency of closed-loop subspace identification with innovation estimation. In *43rd IEEE Conference on Decision and Control*, pages 2195–2200, Atlantis, Bahamas, December 2004.

- [41] B. Liu and J. Si. Fault isolation filter design for linear time-invariant systems. *IEEE Trans. Auto. Cont.*, 42:704–707, 1997.
- [42] L. Ljung. Convergence analysis of parametric identification methods. *IEEE Trans. Auto. Cont.*, 23:770–783, 1978.
- [43] L. Ljung. Asymptotic variance expression for identified black-box transfer function models. *IEEE Trans. Auto. Cont.*, 30:834–844, 1985.
- [44] L. Ljung. Estimation focus in system identification: Prefiltering, noise models, and prediction. In *IEEE Conference on Decision and Control*, pages 2810–2815, Dec 1999.
- [45] L. Ljung. *System Identification - Theory for the User*. Prentice Hall PTR, second edition, 1999.
- [46] L. Ljung and T. McKelvey. Subspace identification from closed loop data. *Signal Processing*, 52:209–215, 1996.
- [47] X. C. Lou, A. S. Willsky, and G. C. Verghese. Optimal robust redundancy relations for failure detection in uncertainty systems. *Automatica*, 22:333–344, 1986.
- [48] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, 1979.
- [49] R. Nikoukhah. Innovations generation in the presence of unknown inputs: application to robust failure detection. *Automatica*, 30:1851–1867, 1994.

- [50] P. Van Overschee and B. De Moor. *Closed-loop subspace system identification*. Technical Report ESAT-SISTA/TR 1996-52I, Katholieke Universiteit Leuven, 1996.
- [51] K. Peternell, W. Scherrer, and M. Deistler. Statistical analysis of novel subspace identification methods. *Signal Processing*, 52:161–177, 1996.
- [52] M. Phan, L. G. Horta, J. N. Juang, and R. W. Longman. Improvement of observer/Kalman filter identification (OKID) by residual whitening. In *AIAA Guidance, Navigation and Control Conference*, Hilton Head, South Carolina, Aug 1992.
- [53] S. J. Qin. Statistical process monitoring: basics and beyond. 17:480–502, 2003.
- [54] S. J. Qin and R. Dunia. Determining the number of principal components for best reconstruction. In *Proc. IFAC DYCOPS'98*, 1998.
- [55] S. J. Qin and W. Li. Detection, identification, and reconstruction of faulty sensors with maximized sensitivity. *AIChE J.*, 45:1963–1976, 1999.
- [56] S. J. Qin and W. Li. Detection and identification of faulty sensors in dynamic processes. *AIChE J.*, 47:1581–1593, 2001.
- [57] S. J. Qin, W. Lin, and L. Ljung. A novel subspace identification approach with enforced causal models. *Revised for Automatica*, 2005.

- [58] S. J. Qin and Lennart Ljung. Closed-loop subspace identification with innovation estimation. In *Proceedings of the 13th IFAC SYSID Symposium*, pages 887–892, Rotterdam, NL, Aug 2003.
- [59] S. J. Qin and Lennart Ljung. Parallel QR implementation of subspace identification with parsimonious models. In *Proceedings of the 13th IFAC SYSID Symposium*, pages 1631–1636, Rotterdam, NL, Aug 2003.
- [60] R. W. Serth and W. A. Heenan. Gross error-detection and data reconciliation in steam-metering systems. *AIChE J.*, 32:733–742, 1986.
- [61] R. Shi and J. F. MacGregor. A framework for subspace identification methods. In *Proceedings of the American Control Conference*, pages 3678–3683, Arlington, VA, June 2001.
- [62] T. Soderstrom and P. Stoica. *System Identification*. Prentice Hall, 1989.
- [63] P. van Overschee and B. de Moor. N4SID: subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30:75–93, 1994.
- [64] P. van Overschee and B. de Moor. A unifying theroem for three subspace system identification algorithms. *Automatica*, 31:1853–1864, 1995.
- [65] P. van Overschee and B. de Moor. *Subspace identification for linear systems*. Dordrecht: Kluwer Academic Publishers, 1996.

- [66] M. Verhaegen. Application of a subspace model identification technique to identify LTI systems operating in closed-loop. *Automatica*, 29:1027–1040, 1993.
- [67] M. Verhaegen. Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. *Automatica*, 30:61–74, 1994.
- [68] M. Verhaegen and P. Dewilde. Subspace model identification. part I: the output-error state-space model identification class of algorithms. *International Journal of Control*, 56:1187–1210, 1992.
- [69] J. Wang and S. J. Qin. A new subspace identification approach based on principle component analysis. *J. Process Control*, 12:841–855, 2002.
- [70] S. Yoon and J. F. MacGregor. Statistical and causal model-based approaches to fault detection and isolation. *AIChE J.*, 46:1813–1824, 2000.

Vita

Weilu Lin was born in Quanzhou, Fujian, P. R. China on 22 March 1972, the youngest son of Zhicheng Lin and Tailing Xu. He received his high school diploma in July, 1990 from Hunan University Affiliated High School in Changsha, Hunan. He entered Xiamen University, Xiamen, and received the Bachelor of Engineering in Chemical Engineering with honor in 1994. In 1997, he received the Master of Engineering in Chemical Engineering from Nanjing University of Chemical Technology, Nanjing. He started working as a research scientist in Process System Engineering Research Center at South China University of Technology, Guangzhou in 1997. In August 2000, Mr. Lin entered the graduate school at the University of Austin and was admitted to Ph.D. candidacy in December, 2003.

Permanent address: 1919 Burton Dr., Apt 252B
Austin, Texas 78741

This dissertation was typeset with L^AT_EX[†] by the author.

[†]L^AT_EX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's T_EX Program.